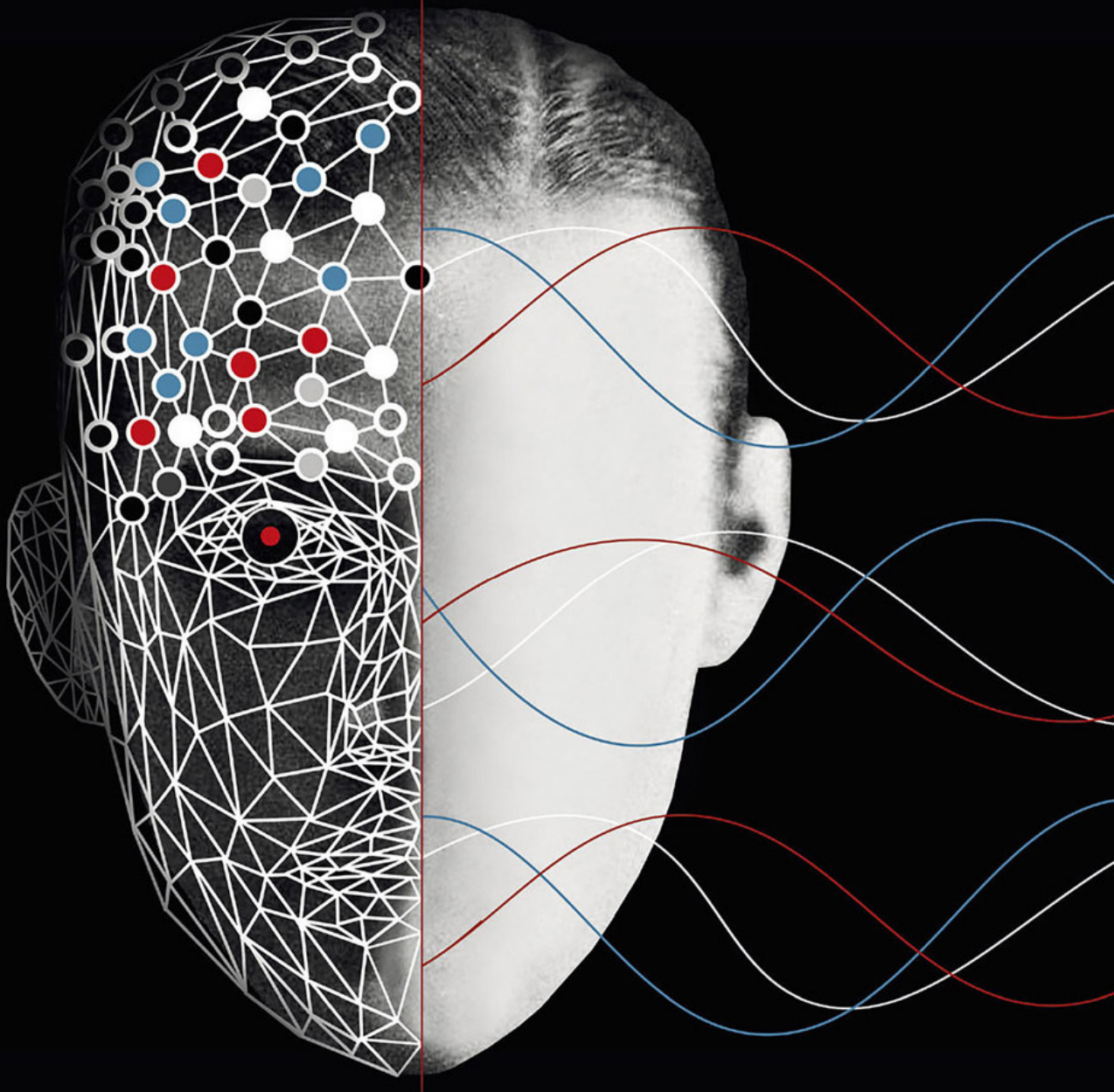
ОТ АВТОМАТОНОВ ДО НЕЙРОСЕТЕЙ

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

ИЛЛЮСТРИРОВАННАЯ ИСТОРИЯ



КЛИФФОРД ПИКОВЕР

Клиффорд Пиковер

Искусственный интеллект. Иллюстрированная история. От автоматов до нейросетей

Издательство «Синдбад»

2019

Пиковер К.

Искусственный интеллект. Иллюстрированная история. От автоматов до нейросетей / К. Пиковер — Издательство «Синдбад», 2019

ISBN 978-5-00131-330-4

Эта книга рассказывает о том, как человечество постепенно, шаг за шагом шло к созданию искусственного интеллекта и как он стал неотъемлемой частью нашей жизни. Как представляли себе искусственный разум ученые, философы и писатели разных эпох? Какие механизмы создавали наши предки, чтобы облегчить решение умственных задач, и в чем сегодня алгоритмы превосходят нас самих? Как сложатся наши отношения с этими новыми сущностями? Что – или кого – мы создаем: верных помощников или потенциальных врагов? С какими этическими проблемами связано использование искусственного интеллекта? Автор не дает окончательных ответов на эти вопросы, но объясняет, в чем их важность, и призывает читателя поразмышлять над ними.

ISBN 978-5-00131-330-4

© Пиковер К., 2019
© Издательство «Синдбад», 2019

Содержание

Предисловие 9
Искусственный интеллект и вокруг него 9
Структура и цель этой книги 11
Крестики-нолики. Ок. 1300 до н. э. 13
Талос. Ок. 400 до н. э. 15
«Органон» Аристотеля. Ок. 350 до н. э. 17
Водяные часы Ктесибия. Ок. 250 до н. э. 19
Счеты. Ок. 190 до н. э. 21
Антикитерский механизм. Ок. 125 до н. э. 23
Автоматы Аль-Джазари. 1206 25
Медные рыцари из легенды о Ланселоте. Ок. 1220 27
Механический парк в Эдене. Ок. 1300 29
Ars Magna Раймунда Луллия. Ок. 1305 31
Религиозные автоматоны. 1352 33
Робот-рыцарь Леонардо да Винчи. Ок. 1495 35
Голем. 1580 37
«Левиафан» Гоббса. 1651 39
Мельница Лейбница. 1714 41
Станок для написания книг в Лагадо. 1726 43
Механическая утка де Вокансона. 1738 45
Механический турок. 1770 47
Автоматоны Жаке-Дро. 1774 49
«Франкенштейн». 1818 51
Машинное творчество. 1821 53
Механический компьютер Бэббиджа. 1822 55
«Мастер красоты». 1844 57
Булева алгебра. 1854 59
«Дарвин среди машин». 1863 61
«Паровой человек в прериях». 1868 63
Ханойская башня. 1883 65
«Большой черный страус Электрического Боба». 1893 67
«Заимствованный разум» Теслы. 1898 69
Тик-ток. 1907 71
Поиски души. 1907 73
Распознавание текста. 1913 75
R.U.R. 1920 77
«Метрополис». 1927 79
Подвижный робот Электро. 1939 81
Синтез речи. 1939 83
Три закона робототехники Азимова. 1942 85
Смертоносные боевые роботы. 1942 87
Искусственные нейронные сети. 1943 89
ЭНИАК. 1946 91
«Гигантские мозги, или Машины, которые думают». 1949 93
Тест Тьюринга. 1950 95
«Человеческое использование человеческих существ». 1950 97

Обучение с подкреплением. 1951 99
Распознавание речи. 1952 101
Обработка естественного языка. 1954 103
Дартмутский семинар по искусственному интеллекту. 1956 105
Перцептрон. 1957 107
Трансгуманизм. 1957 109
Машинное обучение. 1959 111
Представление знаний. 1959 113
«Симбиоз человека и компьютера» Ликлайдера. 1960 115
Психотерапевт ELIZA. 1964 117
Распознавание лиц. 1964 119
Интеллектуальный взрыв. 1965 121
Экспертные системы. 1965 123
Нечеткая логика. 1965 125
Глубокое обучение. 1965 127
Робот Шейки. 1966 129
Жизнь в симуляции. 1967 132
«Кибернетическая проницательность». 1968 134
HAL 9000. 1968 136
Mastermind. 1970 138
«Колосс: проект Форбина». 1970 140
SHRDLU. 1971 142
Параноик PARRY. 1972 144
Генетические алгоритмы. 1975 146
Этика искусственного интеллекта. 1976 148
Победа над чемпионом мира по коротким нардам. 1979 150
Китайская комната. 1980 152
«Бегущий по лезвию». 1982 154
Беспилотные автомобили. 1984 156
«Терминатор». 1984 158
Искусственная жизнь. 1986 160
Роевой интеллект. 1986 162
Парадокс Моравека. 1988 164
Четыре в ряд. 1988 166
«Слоны не играют в шахматы». 1990 168
Герметичный ящик для искусственного интеллекта. 1993 170
Шашки и искусственный интеллект. 1994 172
Тамагочи. 1996 174
Deep Blue обыгрывает чемпиона мира по шахматам. 1997 176
Реверси. 1997 178
Робот AIBO. 1999 180
ASIMO и друзья. 2000 182
«Искусственный разум» Спилберга. 2001 184
Решение для игры вари. 2002 186
Робот-пылесос Roomba. 2002 188
Максимизатор скрепок. 2003 190
Quackle выигрывает в скребл. 2006 192
Watson участвует в телевикторине. 2011 194
Компьютерное искусство и DeepDream. 2015 196

«Назовем их искусственными пришельцами». 2015 198
Искусственный интеллект на Марсе. 2015 200
AlphaGo – чемпион по игре в го. 2016 202
Автономные роботы-хирурги. 2016 204
Искусственный интеллект играет в покер. 2017 206
Состязательные примеры. 2018 208
Роботы собирают кубик Рубика. 2018 210
Предсказатель смерти. 2019 212
Примечания и список литературы 215
Ознакомительное чтение 216
Библиография 217
Иллюстрации 223
Об авторе 224

Клиффорд Пиковер
Искусственный интеллект. Иллюстрированная история. От автоматов до нейросетей

Clifford A. Pickover

ARTIFICIAL INTELLIGENCE: AN ILLUSTRATED HISTORY: FROM MEDIEVAL ROBOTS TO NEURAL NETWORKS

Text © 2019 Clifford A. Pickover

© Издание на русском языке, перевод на русский язык, оформление. Издательство «Синдбад», 2021.

* * *

Мы попытаемся выяснить, как обучить машины использовать естественные языки, формировать абстракции и концепции, решать задачи, сейчас подвластные только людям, и улучшать самих себя… Для заявленной цели проблема искусственного интеллекта заключается в следующем: поведение машины должно быть таким, что, будь это поведение человека, его назвали бы разумным.

Джон Маккарти, Марвин Мински, Натаниэль Рочестер, Клод Шеннон. Заявка на проведение Дартмутского летнего исследовательского проекта по искусственному интеллекту, 1955

Искусственный интеллект способен управлять машинами, торговать акциями на бирже, овладевать сложными навыками, просто просматривая видео на YouTube, переводить с десятков разных языков на другие языки, распознавать лица людей точнее, чем это делаем мы сами, и выдвигать оригинальные гипотезы, помогая нам создавать новые лекарства. И это только начало.

Люк Дормель. Думающие машины, 2017

Только когда машина сможет написать сонет или сочинить концерт благодаря собственным мыслям и эмоциям, а не за счет случайной выдачи символов, мы сможем признать, что эта машина равна мозгу: то есть способна не только написать что-то, но и осознать, что она это написала.

Джеффри Джефферсон. Сознание механического человека, 1949

Основаны ли мы на углероде или на кремнии, не имеет принципиального значения; к каждому из нас следует относиться с должным уважением[1].

Артур Кларк. 2010: Одиссея Два, 1982

Возникнув сразу из многих областей – философии, математики, психологии и даже неврологии, – искусственный интеллект поднимает

[1] Пер. Д. Старкова.

основные вопросы о человеческом интеллекте, памяти, проблеме разума и тела, происхождении языка, символическом мышлении, обработке информации и так далее. Исследователи искусственного интеллекта, подобно алхимикам древности, жаждавшим превратить в золото обычный металл, стремятся создать мыслящую машину из бесконечно малых кусочков оксида кремния.

Даниэль Кревье. Бурная история поиска искусственного интеллекта, 1993

Предисловие

Период существования биологического разума – лишь краткий промежуток между ранними формами жизни и долгой эрой машин.
Мартин Рис. Из интервью сайту The Conversation, апрель 2017.

Искусственный интеллект и вокруг него

Многие новейшие технологии ИИ находят повсеместное применение, часто даже не называясь искусственным интеллектом: как только что-то становится достаточно полезным и распространенным, его перестают называть ИИ.
Ник Бостром. ИИ готовится превзойти возможности человеческого мозга. CNN.com, 2006

На протяжении всей истории человечества тайны разума, природа мышления и возможность создания искусственных существ будоражили воображение художников, ученых, философов и даже богословов. Мифология, изобразительное искусство, музыка и литература полны образов и историй, связанных с *автоматонами* – подвижными механическими устройствами, созданными в подражание живым существам. Наше увлечение искусственным интеллектом (ИИ) – то есть разумным поведением машин – также выражается в жутких и сверхъестественных сюжетах блокбастеров и видеоигр, где появляются роботы, наделенные эмоциями, и существа с совершенным интеллектом, непостижимым для человека.

В этой книге мы отправимся в долгое путешествие во времени от древних игр к современным вычислительным методам, основанным на искусственных нейросетях, которые учатся и совершенствуются, зачастую не требуя – или почти не требуя – целевого программирования и заданных правил. На этом пути нам встретятся диковинные чудеса, например таинственные медные рыцари из цикла легенд о короле Артуре, а также творение французского изобретателя Жака де Вокансона – гиперреалистичная утка-автоматон, которая спустя более 250 лет вдохновила американского писателя Томаса Пинчона на создание одного из персонажей романа «Мейсон и Диксон». Мы познакомимся с каталонским философом XIII в. Раймундом Луллием, который одним из первых системно подошел к вопросу об искусственной генерации идей с помощью механического устройства. Мы перенесемся в 1893 г. и посмотрим на забавного черного страуса, созданного Электрическим Бобом; в этой истории, как и в романе «Паровой человек в прериях», отразился особый интерес ко всему механическому в Викторианскую эпоху – этакий викторианский стимпанк.

Ближе к нашему времени мы встретимся с Артуром Сэмюэлом из *IBM*, который в 1952 г. создал одну из первых компьютерных программ для игры в шашки, а в 1955 г. – программу, которая *научилась* играть в эту игру без посторонней помощи. Сегодня термином «искусственный интеллект» часто обозначают системы, созданные для обучения, решения проблем и взаимодействия с людьми посредством обработки естественного языка. Умные помощники, такие как Алекса от *Amazon*, Сири от *Apple* и Кортана от *Microsoft*, обладают некоторыми возможностями ИИ.

В этой книге мы также рассмотрим непростые этические вопросы, связанные с использованием ИИ, и даже следующую проблему: как поместить сложные системы ИИ – если они станут слишком разумными и опасными – в герметичные ящики, чтобы изолировать их от внешнего мира? Конечно, границы и масштабы ИИ со временем меняются, и некоторые спе-

циалисты предлагают более общие определения, под которые подпадает целый ряд технологий, помогающих людям выполнять мыслительные операции. Поэтому, чтобы шире осветить историю ИИ, я включил в книгу несколько машин и механизмов, которые помогают решать задачи, требующие умственных усилий и подсчетов. Среди таких устройств – счеты, Антикитерский механизм (125 до н. э.), ЭНИАК (1946) и другие изобретения. В конце концов, без этих простейших технологий у нас не было бы современных шахматных программ и систем управления транспортом.

Читая эту книгу, помните: даже если какие-то гипотезы из прошлого или предсказания по поводу искусственных существ кажутся нам неправдоподобными, любая давняя идея внезапно может воплотиться в жизнь, если для этого найдется достаточно быстрый и совершенный компьютер. Наши технологические прогнозы – и даже наши легенды – по меньшей мере представляют собой увлекательную картину познания и творчества и показывают, как мы проникаем в разные культуры и эпохи, чтобы понять друг друга и выяснить, что ценно и сакрально для нашего общества. Но, отдавая дань человеческим воображению и изобретательности, крайне важно задумываться о нежелательных последствиях, в том числе о потенциальной опасности ИИ. В 2014 г. физик-теоретик Стивен Хокинг сказал в интервью Би-би-си: «Развитие полноценного искусственного разума может положить конец человеческой расе… Этот разум возьмет инициативу на себя и станет сам себя совершенствовать со все возрастающей скоростью». Иными словами, существует вероятность, что объекты с ИИ станут настолько разумными и умелыми, что, постоянно улучшая себя, в конце концов создадут некий сверхразум, потенциально чрезвычайно опасный для человечества. Этот сценарий стремительного технологического роста, иногда называемый *технологической сингулярностью* , может привести к невообразимым изменениям в цивилизации, обществе и жизни людей.

Таким образом, несмотря на многочисленные потенциальные преимущества ИИ – беспилотные автомобили, эффективные бизнес-процессы и даже помощь роботов-компаньонов в самых разных делах, – необходимо проявлять особую осторожность при разработке автономных комплексов вооружения и не слишком полагаться на технологии ИИ с порой непостижимыми механизмами. Например, исследования показывают, что некоторые системы распознавания образов на основе ИИ (нейросети) можно легко «обмануть» и заставить ошибочно идентифицировать животных как винтовки или принять самолет за собаку. Для этого достаточно немного изменить изображения таким образом, что люди даже ничего не заподозрят. Если террористу удастся сделать торговый центр или больницу похожими на военную цель для беспилотника, последствия могут быть ужасными. С другой стороны, вполне возможно, что боевые машины с настроенными датчиками и заданными этическими правилами могли бы снизить число жертв среди мирного населения. Чтобы потенциальная опасность ИИ-устройств не перечеркивала их ценные преимущества, в этой сфере необходимо создать продуманную нормативную базу.

Мы все больше полагаемся на ИИ с его многочисленными сложными нейросетями глубокого обучения, и одновременно с этим развивается одна интересная область: разработка систем ИИ, которые смогут *объяснить* человеку, каким образом они пришли к тому или иному решению. Однако, заставив ИИ объяснять самого себя, мы тем самым ограничим его возможности – по крайней мере, в некоторых случаях. Дело в том, что многие из этих систем способны создавать гораздо более сложные модели реальности, чем люди могут себе представить. Эксперт по ИИ Дэвид Ганнинг даже предполагает, что самая высокопроизводительная система окажется и самой труднообъяснимой.

Структура и цель этой книги

Меня давно увлекает вычислительная техника и интересуют проблемы, возникающие на переднем крае науки. В этой книге, адресованной широкой аудитории, я предлагаю краткий путеводитель по любопытным и вместе с тем важным практическим идеям из истории *искусственного интеллекта* – хотя сам этот термин был предложен только в 1955 г. информатиком Джоном Маккарти. Каждая глава состоит всего из нескольких абзацев, так что книгу можно читать с любого места, не продираясь через многословные описания. Конечно, такой формат не позволил мне углубляться в подробности, однако в разделе «Примечания и список литературы» я предлагаю материалы для дальнейшего чтения и поиска источников цитат или трудов упомянутых авторов.

Главы этой книги охватывают такие области, как философия, поп-культура, информатика, социология и теология, а также темы, которые интересуют меня лично. В молодости я был очарован книгой Ясии Рейхардт «Кибернетическая проницательность: компьютер и искусство», опубликованной в 1968 г. В книге рассказывалось, как компьютеры создают стихи, картины, музыку и многое другое. Также меня поражает, каких успехов в области искусства достигли специалисты по ИИ, используя порождающие состязательные сети для создания потрясающих фотореалистичных изображений смоделированных лиц, цветов и птиц. Порождающие состязательные сети – это две нейросети, которые «натравлены» друг на друга: одна генерирует идеи и паттерны, а другая оценивает результаты.

Сегодня возможности применения ИИ кажутся безграничными, и в разработки в этой области ежегодно вкладываются миллиарды долларов. Технологии ИИ использовали, например, для расшифровки документов из Ватиканского секретного архива: ученые пытались разобрать сложные рукописные тексты из этой огромной исторической коллекции. ИИ также помогает прогнозировать землетрясения, интерпретировать медицинские снимки, распознавать речь и предсказывать время смерти пациента на основе информации из его электронной медицинской карты. С помощью ИИ придумывают шутки, игры и головоломки, формулируют математические теоремы, создают патентуемые изобретения, разрабатывают инновационные конструкции антенн, новые оттенки красок, парфюмерные ароматы и многое другое. Уже сегодня многие из нас разговаривают со своими смартфонами и прочими устройствами, а в будущем наши отношения с машинами станут еще более близкими и похожими на человеческие.

Главы расположены в хронологическом порядке с указанием года, связанного с важным событием, книгой или открытием. Датировка часто условна, некоторые годы приводятся приблизительно; там, где это было возможно, я попытался объяснить, почему указал ту или иную дату.

Как легко заметить, больше половины глав приходится на период после 1950 г. Даниэль Кревье, автор книги «Бурная история поиска искусственного интеллекта» (1993), отмечает, что в 1960-е гг. «искусственный интеллект расцвел тысячей цветов. Специалисты по ИИ использовали новые методы программирования для решения многих проблем, которые, хоть и были реальными, оказались значительно упрощены – отчасти ради разделения задач, требующих решения, отчасти для того, чтобы втиснуть их в крошечную память компьютеров того времени».

Тайны сознания, недостатки ИИ и природа разума будут изучаться еще многие годы; но эти проблемы интересовали людей с древних времен. Памела Маккордак в своей книге «Машины, которые думают» высказывает предположение, что ИИ начался с желания древних людей «выковать богов».

Грядущие открытия, связанные с ИИ, войдут в число величайших достижений человечества. История ИИ – это история не только о том, как мы создаем свое будущее, но и о том,

как люди будут жить в условиях бурного развития интеллекта и творческих возможностей. Что будет вкладываться в понятие «человек» через сто лет? Каким будет общество, в котором повсеместно станут использоваться устройства с ИИ? Что произойдет с рабочими местами? Будем ли мы влюбляться в роботов?

Если методы и модели ИИ уже помогают решать, кого нанять на работу, с кем пойти на свидание, кто получит условно-досрочное освобождение, кто более склонен к психическим расстройствам и как управлять беспилотными автомобилями и дронами, то какой уровень контроля над нашей жизнью мы доверим системам ИИ будущего? Если они все чаще принимают решения за нас, легко ли будет обмануть какой-нибудь модуль ИИ и заставить его совершить катастрофическую ошибку? Как специалистам по ИИ выяснить, почему одни алгоритмы и архитектуры машинного обучения эффективнее других, и в то же время упростить воспроизведение чужих экспериментов и их результатов?

Наконец, есть ли гарантии, что устройства на основе ИИ будут действовать этично или у машин когда-либо появятся психические состояния и чувства, свойственные людям? Несомненно, устройства с ИИ станут чем-то вроде протезов для нашего слабого мозга и помогут нам мыслить и мечтать по-новому. Для меня искусственный интеллект – источник постоянного удивления по поводу границ разума, будущего человечества и нашего места в огромном пространственно-временном ландшафте, который мы называем своим домом.

Крестики-нолики. Ок. 1300 до н. э.

По данным археологов, нечто похожее на игру с выстраиванием трех элементов в ряд существовало еще примерно в 1300 г. до н. э. в Древнем Египте. При игре в крестики-нолики два игрока по очереди вписывают свои символы (O или X) в клетки поля размером 3 × 3. Выигрывает тот, кто первым проставит три своих знака в ряд по горизонтали, вертикали или диагонали.

Крестики-нолики попали в эту книгу потому, что их часто упоминают при объяснении базовых принципов программирования и искусственного интеллекта из-за простоты их игровых деревьев (где узлы графа – это позиции в игре, а ребра – ходы). Крестики-нолики – это так называемая игра с полной информацией, поскольку все игроки в курсе всех сделанных ходов. Кроме того, это последовательная игра без рандомизации: игроки ходят по очереди и не используют игральные кости.

Крестики-нолики можно назвать атомом, на основе которого веками формировались молекулы более сложных позиционных игр. Даже при минимальных вариациях и расширениях эта простая игра становится труднейшей задачей, решение которой требует большого количества времени. Математики и любители головоломок усложняли крестики-нолики, добавляя дополнительные клетки и измерения, а также необычные игровые поверхности, например прямоугольные или квадратные поля, соединенные по краям в форме тора (бублика) или бутылки Клейна (поверхности, у которой только одна сторона).

Рассмотрим некоторые любопытные особенности этой игры. Всего существует 362 880 (9! то есть 1 × 2 × 3 × 4 ×… × 9) возможных сценариев заполнения поля двумя игроками. Однако, если рассматривать все возможные партии, при которых игра заканчивается в 5, 6, 7, 8 или 9 ходов, наберется 255 168 таких партий. В 1960 г. ИИ-система *MENACE* (хитроумная конструкция из спичечных коробков и разноцветных шариков) научилась играть в крестики-нолики путем обучения с подкреплением. В начале 1980-х г. компьютерные гении Дэнни Хиллис и Брайан Сильверман с несколькими друзьями сконструировали из 10 тысяч деталей конструктора *Tinkertoy*® компьютер, который играл в крестики-нолики. В 1998 г. ученые и студенты Университета Торонто создали робота для игры в трехмерные крестики-нолики (4 × 4 × 4) с человеком.

СМ. ТАКЖЕ Мельница Лейбница (1714), Обучение с подкреплением (1951), Четыре в ряд (1988), Реверси (1997), Решение для игры вари (2002)

Крестики-нолики можно сделать более сложными для людей и машин с ИИ, расширив стандартное поле 3 × 3 до больших размеров, добавив новые измерения и эффект гравитации, при котором каждый элемент опускается в нижнюю свободную позицию, например как в этой трехмерной версии 4 × 4 × 4.

Талос. Ок. 400 до н. э.

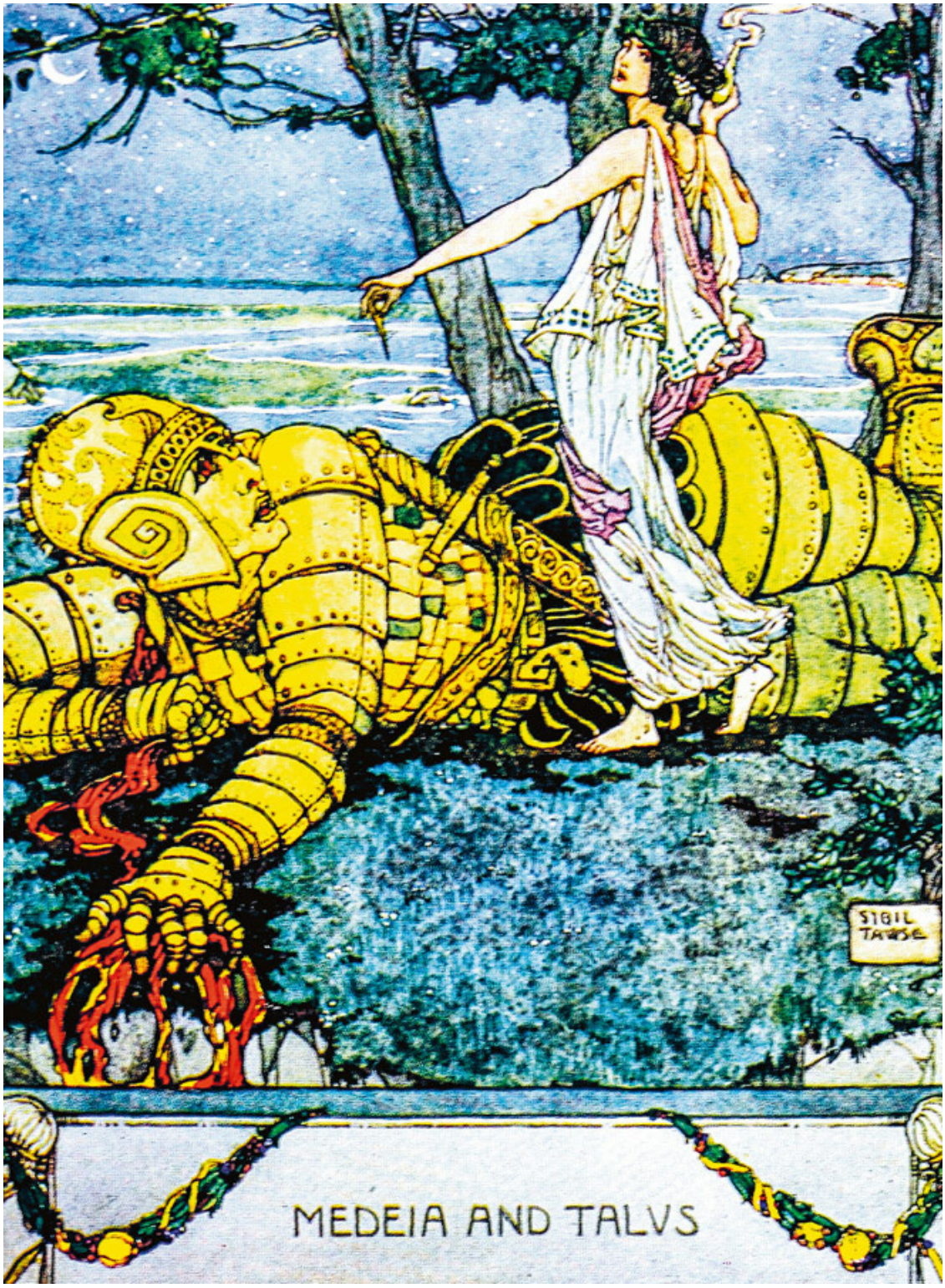
«Многим людям образ Талоса знаком по его воплощению в виде бронзового гиганта в фильме 1963 г. „Ясон и Аргонавты“, – пишет Брайан Хотон. – Но откуда взялась идея Талоса и мог ли он быть первым роботом в истории?»

Согласно греческой мифологии, Талос был огромным бронзовым автоматоном («роботом»), созданным для защиты Европы – матери критского царя Миноса – от захватчиков, пиратов и других врагов. Он был запрограммирован патрулировать берега острова и трижды в день обходил по кругу весь Крит. Порой, чтобы остановить неприятелей, он бросал в них огромные валуны. В других случаях этот гигантский робот прыгал в огонь, раскалялся докрасна, а затем обхватывал тело врага и сжигал его заживо. Иногда Талоса изображали в виде крылатого существа – как на монетах из критского города Феста, датируемых приблизительно 300 г. до н. э. Изображения Талоса также были обнаружены на вазах, созданных около 400 г. до н. э.

Существуют разные версии сотворения и гибели Талоса. В одном мифе его по просьбе Зевса создал Гефест – греческий бог огня и обработки металлов, покровитель кузнецов и других ремесленников. Поскольку Талос был автоматоном, его внутренняя структура по сложности уступала человеческой; по сути, у Талоса имелась одна-единственная вена, которая тянулась от шеи к лодыжке. Снизу вена была запечатана и защищена от протечки бронзовым гвоздем. По одной из легенд, колдунья Медея свела Талоса с ума при помощи духов смерти (их называли «керами») и заставила выбить гвоздь. Божественная кровь (ихор) хлынула у него из лодыжки, «как расплавленный свинец», и великан умер.

Талос – лишь один из примеров того, как древние греки представляли себе роботов и самодвижущиеся автоматы. Здесь также стоит упомянуть труды математика Архита Тарентского (428–347 до н. э.), который, возможно, придумал и создал механического летающего голубя, приводимого в движение паром.

СМ. ТАКЖЕ Водяные часы Ктесибия (ок. 250 до н. э.), Медные рыцари из легенды о Ланселоте (ок. 1220), Голем (1580), «Франкенштейн» (1818)



Изображение Талоса из «Историй о богах и героях» Томаса Булфинча (1920), выполненное английской художницей Сибил Таус (1886–1971).

«Органон» Аристотеля. Ок. 350 до н. э.

Греческий философ Аристотель (384–322 до н. э.) затрагивал в своих работах несколько важных тем, которые и по сей день интересуют исследователей ИИ. В своей книге «Политика» Аристотель высказал предположение, что когда-нибудь автоматы заменят рабов: «Если бы каждое орудие могло выполнять свойственную ему работу само, по данному ему приказанию или даже его предвосхищая, и уподоблялось бы статуям Дедала или треножникам Гефеста, о которых Гомер говорит, что они “сами собой входили в собрание богов”, если бы ткацкие челноки сами ткали, а плектры сами играли на кифаре, тогда и зодчие не нуждались бы в работниках, а господам не нужны были бы рабы»[2].

Аристотель также положил начало системному изучению логики. В своих трудах под общим названием «Органон» (*др.-греч.* «инструмент», «метод») он предлагает приемы выяснения истины и осмысления мира. Основной инструмент в арсенале Аристотеля – *силлогизм*, трехступенчатый аргумент, например: «Все женщины смертны; Клеопатра – женщина; следовательно, Клеопатра смертна». Если две предпосылки истинны, то и заключение должно быть истинным. Аристотель также проводит различие между частностями и универсалиями (то есть общими категориями). Например, *Клеопатра* – это частное понятие, тогда как *женщина* и *смертны* – универсальные. Когда речь идет об универсалиях, им предшествуют слова *все, некоторые* или *ни один*. Аристотель проанализировал множество возможных типов силлогизмов и показал, какие из них состоятельны.

Аристотель также анализировал силлогизмы с модальной логикой – то есть утверждения, содержащие слова *возможно* или *обязательно*. Современная математическая логика далеко ушла от аристотелевской методологии, а его приемы были доработаны для применения к суждениям с другой структурой, включая те, что выражают более сложные отношения, и те, что содержат более одного квантора, как, например, фраза «Ни одному человеку не нравятся все люди, которым не нравятся некоторые люди». И все же глубокие изыскания Аристотеля в области логики считаются одним из величайших достижений человечества, давшим толчок многим разработкам в области математики и искусственного интеллекта.

СМ. ТАКЖЕ Талос (ок. 400 до н. э.), Булева алгебра (1854), Нечеткая логика (1965)

[2] Пер. С. А. Жебелева.

Этот впечатляющий бюст Аристотеля – римская копия бронзового оригинала работы древнегреческого скульптора Лисиппа, жившего в IV в. до н. э.

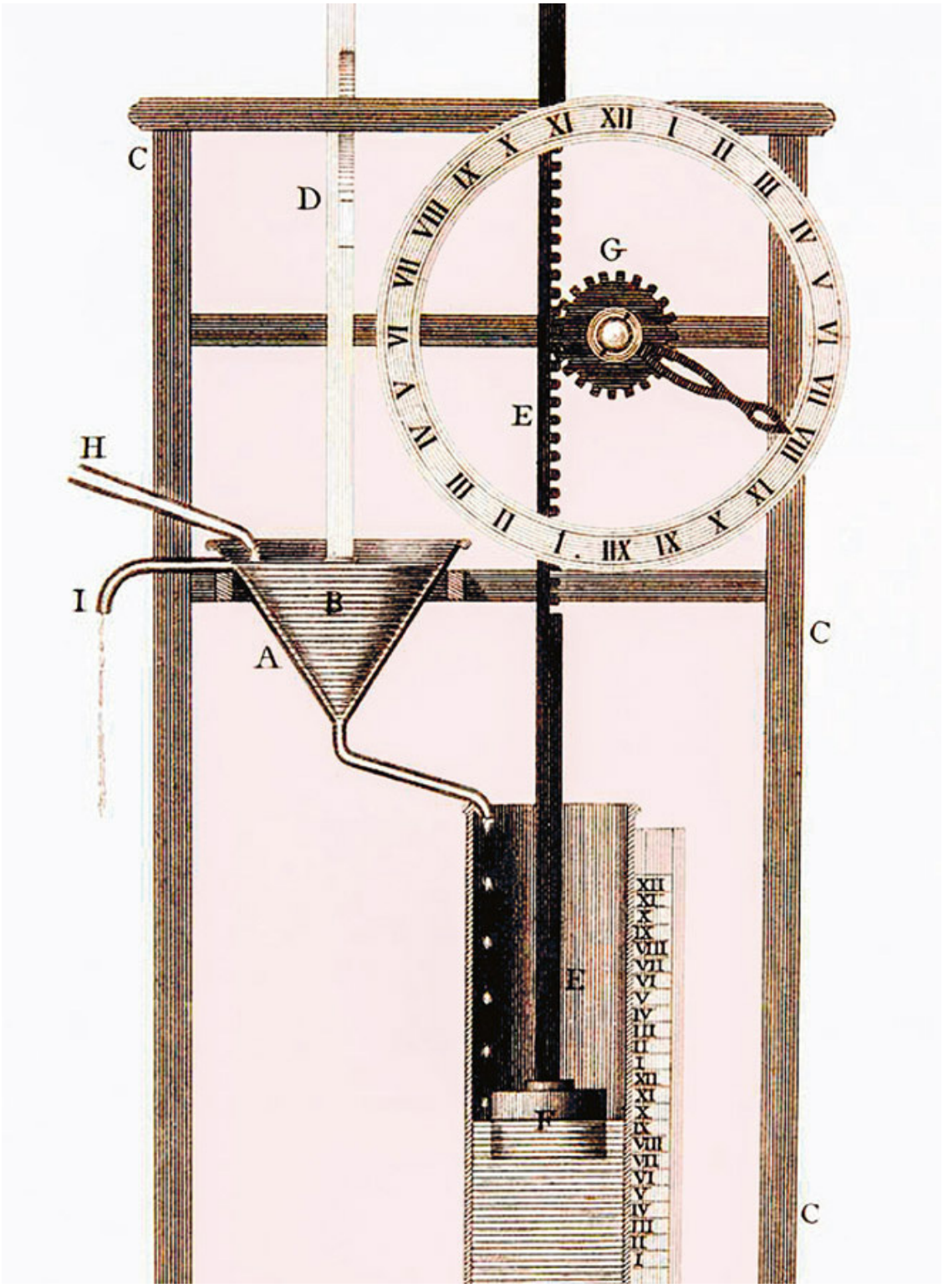
Водяные часы Ктесибия. Ок. 250 до н. э.

«Водяные часы Ктесибия навсегда изменили представление людей о том, на что способен рукотворный объект, – пишет журналист Люк Дормель. – До появления часов Ктесибия считалось, что только живое существо может менять свое поведение с учетом изменений среды. После их изобретения саморегулирующиеся автоматические системы с замкнутым контуром управления стали частью наших технологий».

Греческий изобретатель Ктесибий, или Тесибий (работал в 285–222 гг. до н. э.), прославился в Александрии Египетской благодаря своим устройствам, в том числе насосам и гидравлическим механизмам. Его водяные часы, или *клепсидра* (букв. «похищающая воду»), представляют особый интерес, поскольку в них использовался автоматический регулятор в виде поплавка, который поддерживал постоянный расход воды. Благодаря этому часы довольно точно отмеряли время по уровню воды в приемном сосуде. В одной из версий часов временн*а*я шкала была нанесена на вертикальную стойку, и фигурка в виде человека указывала на ее отметки, поднимаясь вместе с уровнем воды в резервуаре. По некоторым данным, фигурка дополнялась другими механизмами, такими как поворачивающиеся стержни и падающие камни или яйца, а также издавала трубные звуки. С помощью клепсидр Ктесибия отмеряли время, отводимое ораторам на судебных заседаниях, и ограничивали время пребывания посетителей в афинских публичных домах.

Ктесибий, вероятно, был первым руководителем Александрийского мусейона – учреждения, в состав которого входила Александрийская библиотека и которое привлекало ведущих ученых эллинистического мира. Хотя Ктесибий известен как изобретатель особых видов клепсидр, похожие водяные часы создавались в Древнем Китае, Индии, Вавилоне, Египте, Персии и прочих местах. По некоторым данным, Ктесибий также изобрел причудливую роботизированную статую божества, которая использовалась в шествиях (например, в знаменитом Великом параде Птолемея Филадельфа). Этот автоматон умел вставать и садиться благодаря кулачковому механизму (некруглым колесам, преобразующим вращательное движение в линейное), который, возможно, приводился в действие при движении повозки.

СМ. ТАКЖЕ Автоматы Аль-Джазари (1206), Механический парк в Эдене (ок. 1300), Робот-рыцарь Леонардо да Винчи (ок. 1495)



Водяные часы, схема которых приведена на иллюстрации, не обладают всеми функциями часов Ктесибия, но чертеж дает представление о том, как работают подобные устройства. Изображение взято из «Циклопедии, или Универсального словаря искусств, наук и литературы» Абрахама Риса (1820).

Счеты. Ок. 190 до н. э.

«Искусственный интеллект начинался с календаря и счетов, – утверждает инженер и писатель Джефф Криммель. – Искусственный интеллект – это любая технология, которая помогает человеку выполнять умственную задачу. В этом смысле календарь – тоже искусственный интеллект: он дополняет или заменяет нашу память. Точно так же и счеты – искусственный интеллект: они избавляют нас от необходимости выполнять сложные арифметические вычисления в уме».

По разным свидетельствам, устройства для выполнения подсчетов существовали еще в Древней Месопотамии и Египте, но самая старая сохранившаяся счетная доска, или абак, датируется примерно 300 г. до н. э. Это найденная на греческом острове «саламинская доска» – мраморная плита с несколькими группами параллельных желобков. В древности абаки обычно изготавливались из дерева, металла или камня. На дощечки наносились линии или желобки, по которым перекатывались камешки или бусины.

Около 1000 г. н. э. ацтеки изобрели *непоуальцинцин* (его иногда называют ацтекским компьютером) – нечто вроде счетов в виде деревянной рамы с кукурузными зернами, нанизанными на нити. Современные счеты с костяшками, движущимися по спицам, появились не позднее 190 г. в Китае, где их называли *суаньпань*. В Японии такие счеты называются *соробан*.

В некотором смысле счеты можно назвать прообразом компьютера: как и компьютер, они позволяют быстро выполнять вычисления, связанные с торговыми операциями и техническими задачами. Претерпев небольшие изменения в конструкции, счеты и по сей день используются в Китае, Японии, странах Африки и бывшего СССР. Хотя обычно счеты применяют для сложения и вычитания, опытные пользователи умеют с их помощью быстро умножать, делить и извлекать квадратные корни. В 1946 г. в Токио было проведено соревнование по вычислениям между японским оператором соробана и пользователем электронного калькулятора, чтобы выяснить, какой инструмент позволяет быстрее решить некоторые арифметические задачи. В большинстве случаев человек с соробаном опережал электронный калькулятор.

Счеты сыграли столь важную роль в истории, что в 2005 г. читатели, редакторы и эксперты *Forbes.com* назвали их вторым по значимости инструментом всех времен с точки зрения влияния на человеческую цивилизацию. Первую и третью позиции в списке заняли соответственно нож и компас.

СМ. ТАКЖЕ Антикитерский механизм (ок. 125 до н. э.), Механический компьютер Бэббиджа (1822), ЭНИАК (1946)

Счеты оказали огромное влияние на человеческую цивилизацию. На протяжении многих веков они помогали людям быстро выполнять торговые и инженерные расчеты. Европейцы начали использовать счеты задолго до того, как перешли на индо-арабскую систему счисления.

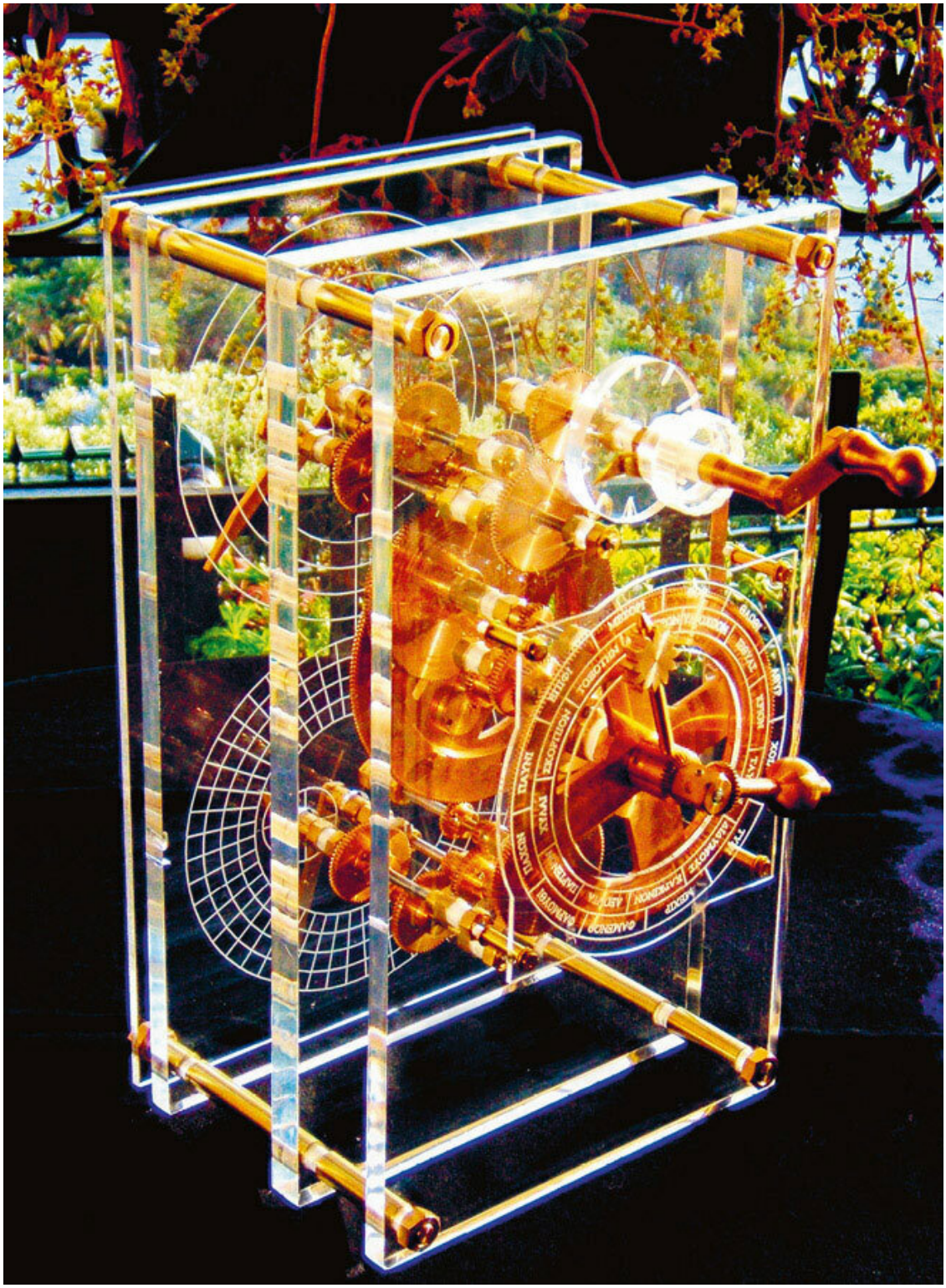
Антикитерский механизм. Ок. 125 до н. э.

В своей книге «Искусственный интеллект» психолог Алан Гарнэм, говоря об антикитерском механизме, отмечает: «Пожалуй, главным направлением развития, обусловившим появление ИИ, были попытки создать машины, которые избавили бы людей от монотонного умственного труда и в то же время устранили некоторые человеческие ошибки». Антикитерский механизм – это древнее вычислительное устройство с зубчатой передачей, которое использовалось для астрономических расчетов. Примерно в 1900 г. механизм обнаружили водолазы у берегов греческого острова Антикитера среди останков затонувшего древнего судна. Предположительно он был создан около 150–100 г. до н. э. Журналистка Джо Марчант рассказывает: «Среди спасенных сокровищ, отправленных в Афины, обнаружился бесформенный булыжник, на который поначалу никто не обращал внимания. Но затем он треснул, обнажив бронзовые шестеренки, стрелки и мелкие надписи на древнегреческом. Хитроумный механизм состоял из ювелирно выгравированных циферблатов, стрелок и по меньшей мере 30 соединенных между собой шестеренок. Ничего близкого по сложности не упоминалось в исторических источниках более тысячи лет, вплоть до появления астрономических часов в средневековой Европе».

На циферблате с передней стороны устройства, судя по всему, было по меньшей мере три стрелки: одна указывала на дату, а две другие – на положение Солнца и Луны. Вероятно, прибор также использовался для определения даты Олимпийских игр, предсказания солнечных затмений и определения движения планет.

В лунном механизме использовалась особая система бронзовых шестеренок, две из которых соединялись со слегка смещенной осью для определения положения и фазы Луны. Как мы знаем сегодня из законов движения планет Кеплера, Луна, вращаясь вокруг Земли, движется с разной скоростью (чем ближе к Земле, тем быстрее). Антикитерский механизм воспроизводил эту разницу – хотя древние греки не знали о том, что лунная орбита имеет форму эллипса. Марчант пишет: «Поворачивая ручку на коробке, можно было заставить время двигаться вперед или назад, чтобы увидеть состояние космоса сегодня, завтра, в прошлый вторник или через сто лет. Тот, кто владел этим устройством, должно быть, чувствовал себя властелином небес».

СМ. ТАКЖЕ Водяные часы Ктесибия (ок. 250 до н. э.), Счеты (ок. 190 до н. э.), Механический компьютер Бэббиджа (1822)



Современная реконструкция антикитерского механизма с воспроизведением рычагов и шестеренок.

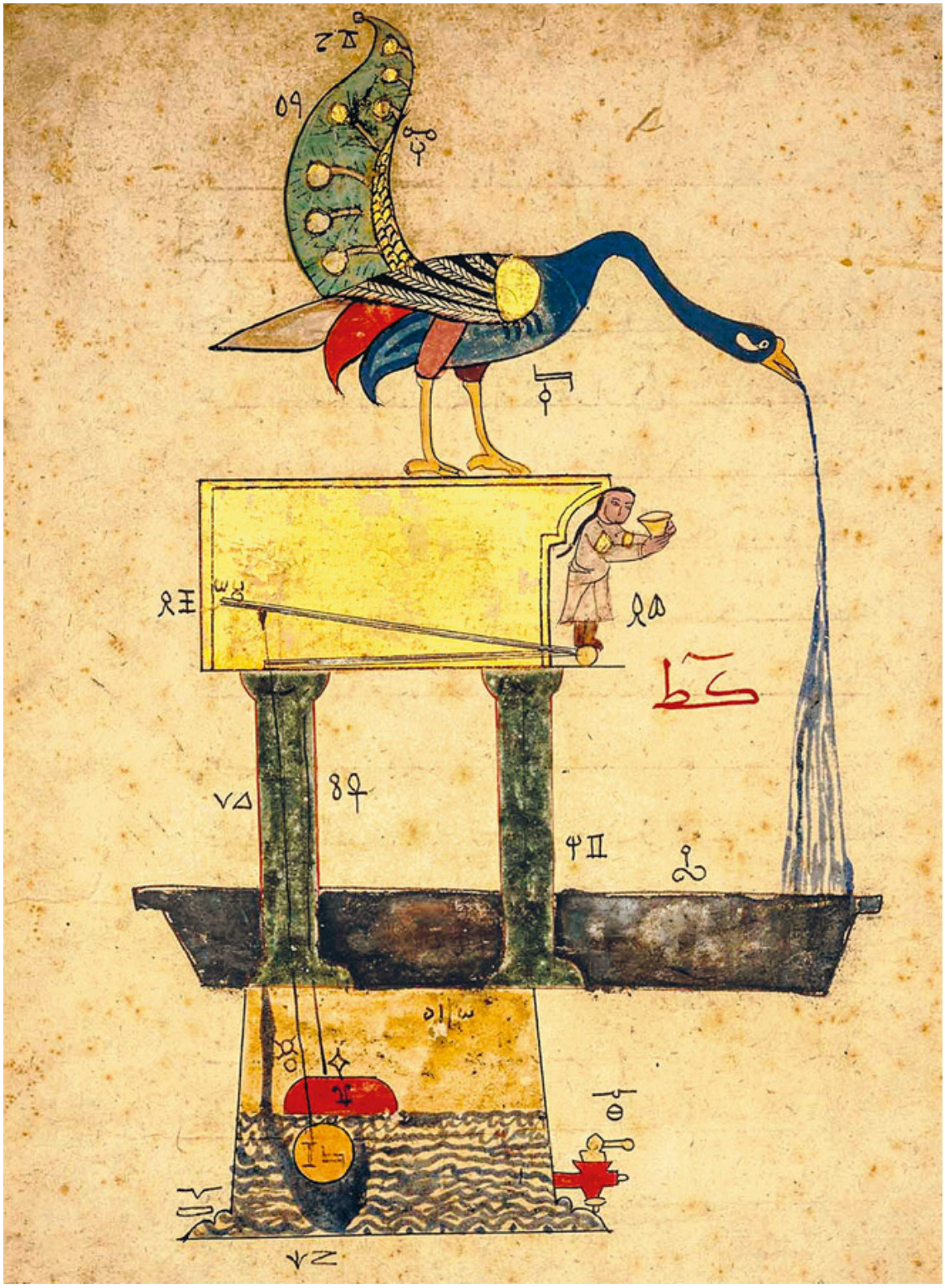
Автоматы Аль-Джазари. 1206

Эрудит, изобретатель, художник и инженер Исмаил Аль-Джазари (1136–1206) жил в эпоху Исламского возрождения. Он сменил своего отца на посту главного инженера во дворце Артуклу в Анатолии (на территории современного турецкого города Диярбакыр). Его трактат «Книга знаний об остроумных механических устройствах», написанный по просьбе правителя и увидевший свет в год смерти автора, содержит описания разнообразных механических устройств, созданных изобретателем, включая движущиеся автоматы в виде людей и животных (автоматоны), а также водоподъемные машины, фонтаны и часы. В процессе исследований и конструирования механизмов Аль-Джазари использовал распределительные и коленчатые валы, храповые механизмы и шестеренки, а также другие сложные устройства.

Среди его автоматов – павлины, приводимые в движение водой, «служанка», подающая напитки, и квартет из механических музыкантов, сидящих в лодке. За счет вращающихся стержней у роботов-музыкантов даже менялись выражения лиц. По предположению некоторых исследователей, движения этих роботов могли быть программируемыми, что говорит о высокой технической сложности механизмов. Конструкция «слоновых часов» Аль-Джазари включала в себя человекоподобного робота, бьющего по тарелке через равные промежутки времени, а также механическую птицу, которая щебетала, когда другой робот – писец – поворачивался, отмечая время своим пером. «Замковые часы» высотой 3,4 метра украшали пять механических музыкантов.

По словам английского инженера и историка Дональда Хилла (1922–1994), который перевел трактат изобретателя на английский язык, «невозможно переоценить значение трудов Аль-Джазари для истории техники. До начала Нового времени ни в одной культуре не появлялось другого такого документа, который содержал бы сопоставимое богатство инструкций по проектированию, изготовлению и сборке машин. Конечно, отчасти это объясняется социальной и культурной пропастью между теми, кто конструировал устройства, и теми, кто писал трактаты. Описывая машину, созданную необразованным мастером, ученый обычно уделял внимание конечному продукту; он не разбирался и не пытался разобраться в хлопотном деле строительства… Поэтому мы в большом долгу перед [работодателем Аль-Джазари] за возможность обладать уникальным документом».

СМ. ТАКЖЕ Водяные часы Ктесибия (ок. 250 до н. э.), Механический парк в Эдене (ок. 1300), Религиозные автоматоны (1352), Автоматоны Жаке-Дро (1774)



Замысловатая чаша с павлином из трактата Аль-Джазари «Книга знаний об остроумных механических устройствах» (бумага, непрозрачная акварель, золото и чернила).

Медные рыцари из легенды о Ланселоте. Ок. 1220

Простые образцы ИИ – механические люди и прочие существа – пользовались популярностью в средневековой Европе. Историк Элли Труитт пишет: «Золотые звери и птицы, музыкальные фонтаны и роботы-слуги поражали и пугали гостей… Автоматы возникали на пересечении естественных знаний (включая магические представления) и технологий, а границы между искусством и природой порой стирались». Реальные и вымышленные устройства, описанные в литературе того времени, составляют любопытную картину «взаимовлияния науки, технологий и воображения».

Вымышленные роботы упоминаются и в «Книге о Ланселоте Озерном» (ок. 1220) – старофранцузском повествовании о приключениях короля Артура и рыцарей Круглого стола, а также о тайном романе рыцаря Ланселота и Гвиневры, жены Артура. Под стенами страшного заколдованного замка Долороз-Гард Ланселот сражается с небольшой армией медных рыцарей-роботов. Проникнув в замок, он побеждает еще двух вооруженных мечами медных рыцарей, охраняющих комнату, где медная девушка держит ключи от заклятия. С помощью ключей Ланселот отпирает ящик, содержащий 30 медных трубок, из которых доносятся ужасающие звуки, и быстро засыпает. Проснувшись, он видит, что медная девушка лежит на полу, а рыцари-роботы развалились на куски.

Историк Джессика Рискин пишет: «Наряду с механическими рыцарями и девами из легенды о короле Артуре в литературе упоминались золотые, серебряные и медные дети, сатиры, лучники, музыканты, оракулы и великаны. У этих вымышленных искусственных существ было множество реальных аналогов. Европу позднего Средневековья и начала Нового времени заполняли настоящие механические люди и животные». Примерно в то же время, когда появилась история о Ланселоте и медных рыцарях, французский художник и инженер Виллар де Оннекур (работал ок. 1225 – ок. 1250) сконструировал механического орла, который поворачивал голову к дьякону, когда тот читал Евангелие. Рискин отмечает, что эти реалистичные автоматы создали предпосылки для появления в XVII в. научных и философских концепций, по которым живые существа рассматривались как машины.

СМ. ТАКЖЕ Талос (ок. 400 до н. э.), Механический парк в Эдене (ок. 1300), Робот-рыцарь Леонардо да Винчи (ок. 1495), Голем (1580), Тик-Ток (1907)

Ланселот сражается с медными рыцарями-автоматами, чтобы проникнуть в замок Долороз-Гард. Человекоподобные рыцари часто изображались без одежды. (Иллюстрация из «Книги о Ланселоте Озерном», Франция, XV в. Париж, Национальная библиотека Франции, *MS Fr. 118, fol. 200v.*)

Механический парк в Эдене. Ок. 1300

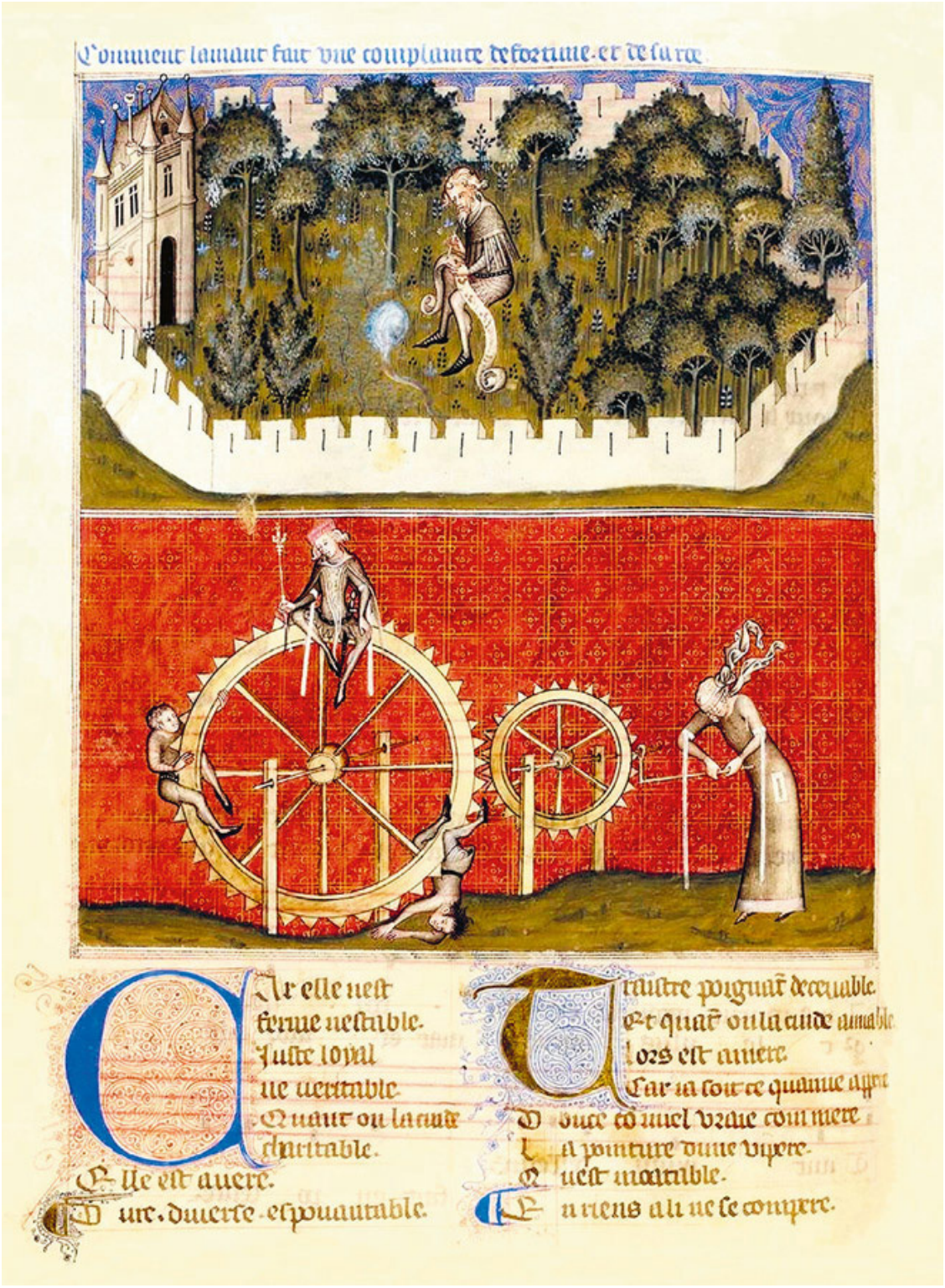
Примерно в 1300 г. парк в Эдене, области на северо-востоке Франции, превратился в знаменитое собрание реалистичных моделей людей и животных. Среди автоматов в Эдене были андроиды, обезьяны, птицы и устройства для отсчета времени. Первые машины в Эдене создавались по заказу Роберта II д'Артуа (1250–1302). В их числе был мост с шестью группами механических обезьян, для большей реалистичности покрытых барсучьей шерстью. Стену павильона украшала механизированная кабанья голова. Когда Роберт умер, его дочь Маго (1268–1329) стала покровительницей новаторских технологий и продолжала поддерживать «механизмы для развлечений» в рабочем состоянии. Так, в 1312 г. обезьян покрыли новым мехом и приделали им рога, чтобы сделать похожими на демонов.

Идея парка автоматов, возможно, была заимствована из исламской культуры – как и вымышленные роботы во французских рыцарских романах. Историк Скотт Лайтси пишет: «Парк в Эдене занимал важное место в европейской традиции искусственных чудес: на его примере можно видеть, как сверхъестественные явления вытеснялись мотивами придворной жизни… Технологические новшества позволяли воспроизводить в изысканных залах и увеселительных садах даже традиционные волшебные образы из рыцарских романов».

В последующие годы в Эденском парке появлялись новые механические диковины, включая деревянного отшельника, который беседовал с посетителями, говорящую сову и фонтаны с механическими птицами. Машущие лапами обезьяны и другие автоматы, скорее всего, управлялись весовыми и заводными механизмами и/или гидравликой.

Эти роботизированные устройства приглашали посетителей заглянуть в будущее, где автоматизация стала повсеместной. Историк Сильвио Бедини отмечает: «Роль автоматов в совершенствовании техники очень велика. Попытки воспроизводить жизнь механическими средствами привели к развитию принципов механики и к созданию сложных машин, которые выполняли изначальную задачу технологий – заменять или упрощать физический труд».

СМ. ТАКЖЕ Водяные часы Ктесибия (ок. 250 до н. э.), Автоматы Аль-Джазари (1206), Медные рыцари из легенды о Ланселоте (ок. 1220), Религиозные автоматоны (1352), Механическая утка де Вокансона (1738), Автоматоны Жаке-Дро (1774)



Парк в Эдене. Изображение обнесенного стеной Эденского сада (вверху). На иллюстрации с шестеренками (внизу) изображена Фортуна, вращающая механизм, который может быть аналогом одного из автоматов в Эдене. (Из книги Гийома де Машо «Снадобье Фортуны», Франция, ок. 1350–1356. Париж, Национальная библиотека Франции, *MS Fr. 1586, fol. 30v.*)

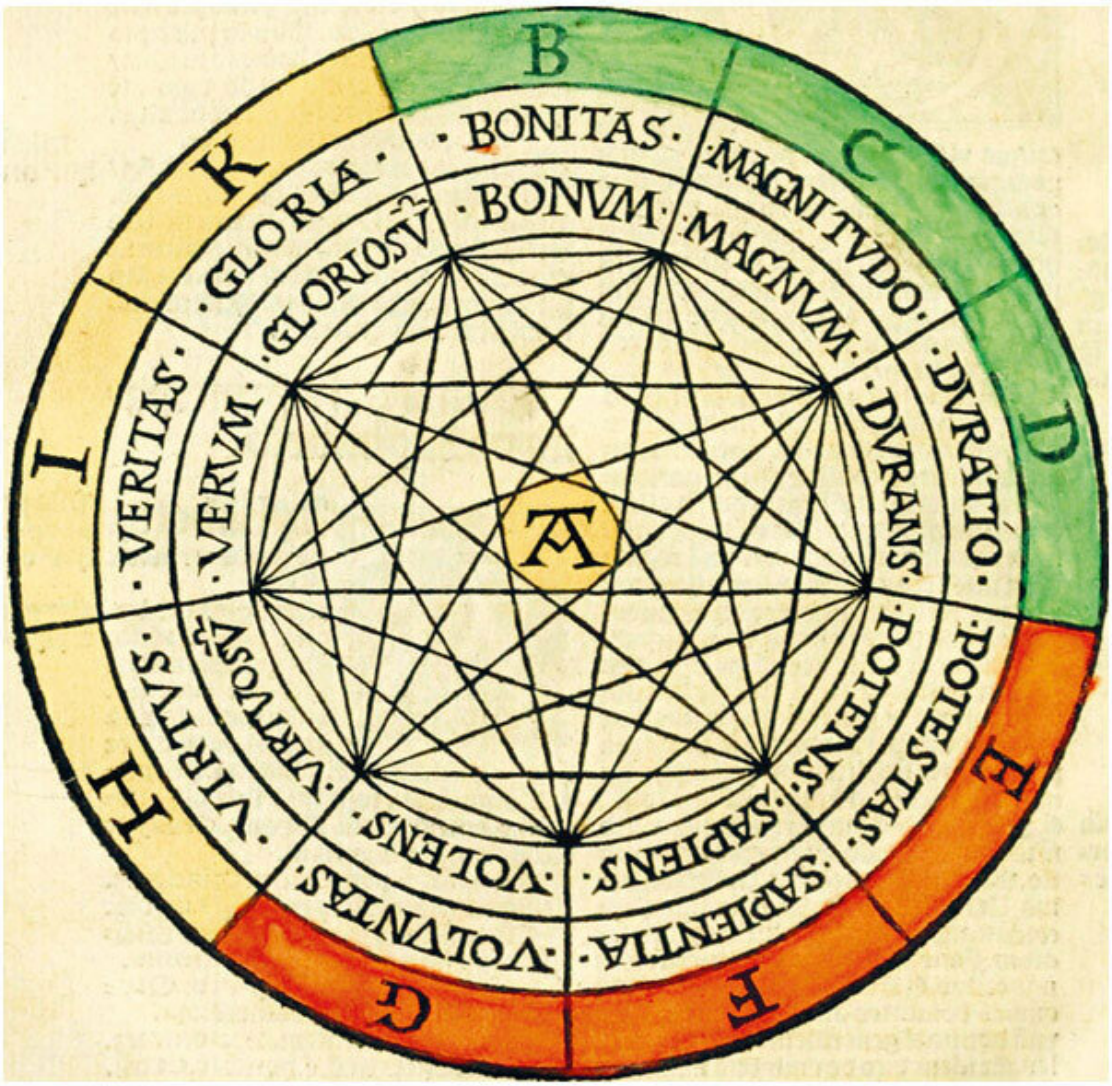
Ars Magna Раймунда Луллия. Ок. 1305

«Поиски искусственного интеллекта, как и все изыскания, начинаются с мечты, – пишет информатик Нильс Нильссон. – Люди издавна фантазировали о машинах с человеческими способностями: подвижных автоматах и мыслящих устройствах». Одно из самых ранних устройств в истории ИИ – луллианский круг. Каталонский философ Раймунд Луллий (или Рамон Льюль, ок. 1232 – ок. 1315) описал в своей книге *Ars Magna* («Великая наука», ок. 1305) бумажную конструкцию из вращающихся концентрических кругов с буквами и словами, написанными по периметру. Подобно механизму кодового замка, символы и слова могли выстраиваться в новые неожиданные комбинации, порождая фонтаны новых идей и логических ходов. Мартин Гарднер пишет: «Это была самая ранняя в истории формальной логики попытка использовать векторные диаграммы в поиске нематематических истин и первая попытка использовать механическое устройство – примитивную логическую машину – для упрощения работы логической системы».

По словам Георги Далакова, устройства Луллия для комбинаторного творчества – одна из первых попыток применить «логические средства для производства знания»: «Луллий крайне простым, но действенным способом доказал, что человеческую мысль можно описать и даже имитировать с помощью устройства. Это был маленький шаг к мыслящей машине». Представим себе Луллия, сидящего за столом при свечах и вращающего диски, чтобы соединить слова. Как пишет Кристина Мадей, он верил, что ему «откроется высшее знание и будут даны логические ответы на вопросы о религии и творении… [Он хотел] исследовать истины и производить новые доказательства с помощью этих комбинаторных устройств».

Немецкий энциклопедист и один из основателей математического анализа Готфрид Лейбниц (1646–1716) вдохновлялся работой Луллия, проводя изыскания в области формальной логики и создавая свой знаменитый арифмометр. Специалист по данным профессор Джонатан Грей пишет: «Первые ручейки загадочных комбинаторных фантазий Луллия и Лейбница постепенно уступили место вездесущим вычислительным технологиям, практикам и стандартам, ныне вплетенным в ткань нашей жизни. Мы пока еще не до конца осознаем их влияние… независимо от того, работают ли машины так, как мы себе представляем».

СМ. ТАКЖЕ Станок для написания книг в Лагадо (1726), Машинное творчество (1821), «Кибернетическая проницательность» (1968)



Один из наборов вращающихся дисков и комбинаций из книги *Ars Magna* Раймунда Луллия. (Из *Illuminati sacre pagine p[ro]fessoris amplissimi magistri Raymundi Lull*, 1517.)

Религиозные автоматоны. 1352

В Европе периода позднего Средневековья и раннего Нового времени появилось множество самых разных автоматонов, связанных с христианской церковью, – от подвижных статуй Христа до механических дьяволов и фигур Сатаны, которые издавали звуки и высовывали языки. Например, в XIV в. в английском аббатстве Боксли в Кенте хранился «крест благодати» – механическая фигура распятого на кресте Иисуса, у которой двигались глаза, губы и другие части тела. Ангелы-автоматоны и механизированные сцены библейских событий стали распространенным явлением в конце XV в. Профессор истории Джессика Рискин пишет: «Автоматоны были привычными элементами повседневной жизни. Они появлялись в церквях и соборах и распространялись оттуда повсеместно. Миссионеры-иезуиты везли их в Китай в качестве подношений, чтобы подчеркнуть величие христианской Европы».

Особый интерес представляют астрономические часы в Страсбургском соборе Нотр-Дам в Эльзасе (Франция). Их создание началось в 1352 г. На огромных часах сидел петух-автоматон, который в определенное время крутил головой, хлопал крыльями и кукарекал (для этого использовались воздуходувные меха и вибрирующая пластина). Среди прочего, часы украшали движущиеся фигуры ангелов. Приблизительно в 1547 г. эти часы были заменены новыми, но механический петух сохранился. Вторые часы перестали работать в 1788 г., но только в 1838 г. на их месте появились нынешние – с новыми механизмами взамен старых.

Помимо автоматонов, на страсбургских часах имеется вечный календарь (с механизмом для определения дней Пасхи), устройство для отображения солнечных и лунных затмений и многое другое. В 1896 г. писательница Фэнни Коу сравнивала страсбургские часы с маленьким театром: «Множество людей и животных играют там свои небольшие роли… Ангелы отбивают круглые часы, а в полдень и полночь из дверей выходят ростовые фигуры Христа и двенадцати апостолов… Затем позолоченный петух на башенке часов хлопает крыльями и кукарекает».

Ученый Джулиус Фрейзер отмечает: «Календарная наука и часовое мастерство совершенствовались, позволяя создавать артефакты, объясняющие и прославляющие христианский мир… [Они были] предвестниками будущего стремления поставить труд ученых и ремесленников на службу земному благу людей».

СМ. ТАКЖЕ Автоматы Аль-Джазари (1206), Механический парк в Эдене (ок. 1300), Автоматоны Жаке-Дро (1774)

Астрономические часы в Страсбургском соборе Нотр-Дам в Эльзасе (Франция). В верхнем левом углу – механический петух.

Робот-рыцарь Леонардо да Винчи. Ок. 1495

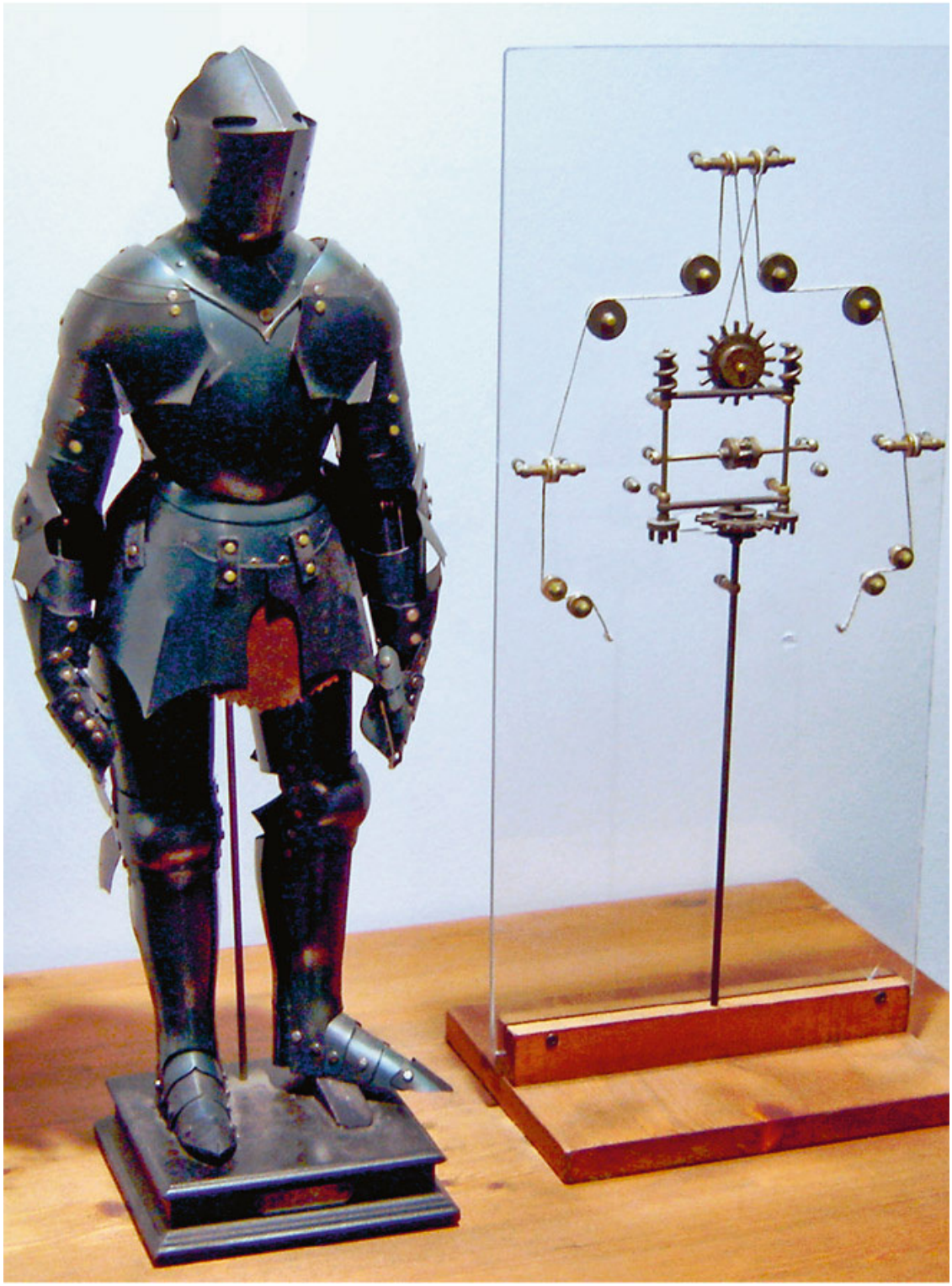
«Закованный в доспехи робот-рыцарь да Винчи мог садиться и вставать, разводить руки и сводить их в хватательном движении; поворачивать голову на гибкой шее и поднимать забрало – возможно, чтобы явить миру жутковатую физиономию. Созданный из дерева, латуни (или бронзы) и кожи, он управлялся с помощью тросов», – пишет инженер-робототехник Марк Розгейм.

Интересы Леонардо да Винчи (1452–1519), итальянского гения эпохи Ренессанса, простирались от живописи и архитектуры до анатомии и инженерного искусства. Среди прочего, в его дневниках можно найти описания и эскизы музыкальных инструментов, кривошипных механизмов и вышеупомянутого механического рыцаря, который был нарисован примерно в 1495 г. Это изображение содержится в «Атлантическом кодексе» (*Codex Atlanticus*) – двенадцатитомном собрании рисунков и сочинений да Винчи. Благодаря шарнирным соединениям у робота-рыцаря были подвижные руки, нижняя челюсть и голова, а внутренний механизм включал систему блоков для приведения их в действие. Ренессансный андроид был облачен в немецко-итальянские средневековые доспехи, мог садиться и вставать и был оснащен несколькими системами зубчатых передач, позволявшими отдельно управлять верхней и нижней частями тела. У Леонардо также встречаются чертежи других автоматов, например механических птиц и повозок.

Хотя мы не знаем, был ли механический рыцарь Леонардо сконструирован на самом деле, подобные роботы могли вдохновлять других изобретателей, таких как итало-испанский инженер Хуанело Турриано (ок. 1500–1585). С помощью системы тросов и блоков Турриано изготовил механическую копию францисканского проповедника Диего для короля Филиппа II Испанского – тот считал, что монах исцелил принца Карлоса от тяжелой травмы головы. Заводной Диего, приводимый в движение пружиной и ключом, ходил, жестикулируя и приоткрывая рот в безмолвной молитве. Сейчас он хранится в Смитсоновском институте в Вашингтоне – и все еще работает.

Размышляя о рыцаре-роботе да Винчи, Синтия Филлипс и Шана Прайвер пишут: «Проект робота Леонардо стал венцом его исследований в области анатомии и геометрии. Как лучше всего объединить механику и человеческое тело? Он взял пропорции и закономерности, характерные для римской архитектуры, и применил их к движениям и поведению, присущим живым существам. В каком-то смысле этот робот был ожившим витрувианским человеком».

СМ. ТАКЖЕ Талос (ок. 400 до н. э.), Медные рыцари из легенды о Ланселоте (ок. 1220), Тик-Ток (1907), Подвижный робот Электро (1939)



Модель робота-рыцаря Леонардо да Винчи, а также его внутренние шестеренки, блоки и тросы.

Голем. 1580

Как писала газета *Forward*, «задолго до того, как Стивен Хокинг предупредил нас об опасности искусственного интеллекта, легенда о големе подспудно сообщала еврейскому народу о том же». Голем из еврейского фольклора – это ожившее существо, вылепленное из глины или грязи. Этот рукотворный «автоматон» демонстрирует несколько форм ИИ, и, как только «технологию активируют» и выпускают на волю, его становится трудно контролировать. Пожалуй, самый знаменитый голем – тот, которого якобы создал в 1580 г. пражский раввин Иехуда Лёв Бен Бецалель (ок. 1520–1609) для защиты евреев Пражского гетто от погромов. История пражского голема была записана несколькими авторами в 1800-х гг.

Обычно голема оживляли с помощью магических или религиозных слов. Согласно легендам, создатели этих существ писали имя Бога у голема на лбу или клали ему под язык глиняную табличку или листок бумаги с тем же словом. Некоторых големов оживляли, нанося им на лоб слово *emet* («истина» на древнееврейском). Стерев первую букву, чтобы получилось слово *met* (на древнееврейском «мертвый»), можно было обездвижить голема.

Согласно некоторым другим старым инструкциям, для создания голема требовалось сопоставить каждую букву еврейского алфавита с каждой буквой тетраграмматона *YHVH* (еврейского имени Бога), а затем произнести каждое из полученных сочетаний со всеми возможными гласными звуками. Тетраграмматон служил «ключом активации», позволяющим пробить брешь в реальности и побудить существо к действиям.

Слово «голем» встречается в Библии только один раз (Псалтирь 139: 16) и обозначает несформировавшееся тело. В современном переводе этот стих звучит так: «Зародыш мой видели очи Твои; в Твоей книге записаны все дни, для меня назначенные, когда ни одного из них еще не было»[3]. На древнееврейском *golem* может обозначать «бесформенную массу» или «неразумную» сущность, а в Талмуде это слово используется в значении «нечто несовершенное». Вот почему в литературе големы обычно изображаются немыми, но им можно поручать простые монотонные задачи. По сути, для создателя главная проблема заключалась в том, чтобы придумать, как в конце концов заставить голема остановиться.

СМ. ТАКЖЕ Талос (ок. 400 до н. э.), Медные рыцари из легенды о Ланселоте (ок. 1220), «Франкенштейн» (1818)

[3] В православной традиции этот псалом фигурирует под номером 138.

Пражский Голем. На картине российско-чешского художника Евгения Иванова изображены голем (крупная фигура в центре) и раввин Лёв Бен Бецалель, сидящий у него на плече.

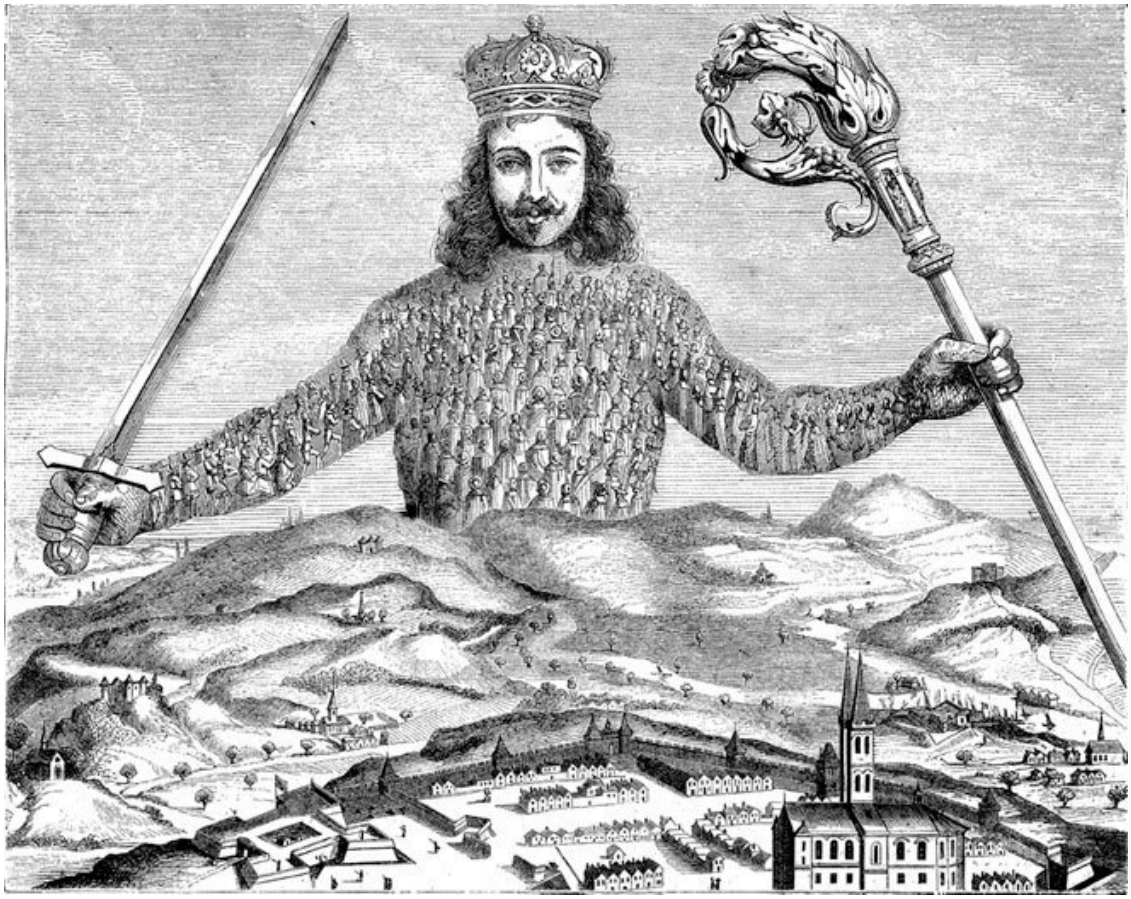
«Левиафан» Гоббса. 1651

В 1651 г. английский философ Томас Гоббс (1588–1679) написал трактат «Левиафан», посвященный структуре общества и его взаимоотношениям с государством. В книге приводится ряд тезисов, которые побудили историка науки Джорджа Дайсона назвать Гоббса патриархом искусственного интеллекта. Например, во введении Гоббс сравнивает тело с механическим двигателем: «Человеческое *искусство* подражает природе (искусству, при помощи которого Бог создал мир и управляет им)… и в том отношении, что человек может создать искусственное животное. Ибо наблюдая, что жизнь есть лишь движение членов, начало которому дается в какой-нибудь основной внутренней части, разве не можем мы сказать, что все *автоматы* (механизмы, движущиеся при помощи пружин и колес, как, например, часы) имеют искусственную жизнь? В самом деле, что такое *сердце*, как не *пружина*? Что такое *нервы*, как не множество *тросов*, а *суставы* – как не множество *колес*, сообщающих движение всему телу?..»

Согласно Гоббсу, когда человек рассуждает, он совершает аналитические вычисления и манипуляции сродни сложению и вычитанию: «Под рассуждением [мышлением] я подразумеваю вычисление. Вычислять – значит либо подсчитывать сумму многих вещей, собранных вместе, либо выяснять, что остается, когда одна вещь отнимается от другой».

Дайсон задается вопросом: «Если размышления можно свести к арифметике, с которой механизмы справлялись еще во времена Гоббса, то способен ли механизм размышлять? Могут ли машины думать?» А вот размышления специалиста по компьютерным архитектурам Дэниела Хиллиса по поводу создания искусственного разума, способного самостоятельно мыслить: «Некоторых пугают механистические интерпретации человеческого разума, и таким людям наше неведение идет во благо: мы не знаем, как простейшие импульсы превращаются в сложные и непредсказуемые системы, и это создает некую успокаивающую дымку. Хотя у отдельных компьютеров и программ развиваются элементы ИИ, Левиафан искусственного разума с большей вероятностью возникнет в создаваемых нами крупных сетях (или в сети как таковой)».

СМ. ТАКЖЕ Мельница Лейбница (1714), «Мастер красоты» (1844), «Дарвин среди машин» (1863), «Гигантские мозги, или Машины, которые думают» (1949), «Человеческое использование человеческих существ» (1950)



Фронтиспис «Левиафана» Томаса Гоббса с гравюрой французского художника Абрахама Босса (1604–1676).

Мельница Лейбница. 1714

Если мы считаем, что сознание – это результат комбинаций и динамических взаимосвязей клеток мозга и их компонентов, то наши мысли, эмоции и воспоминания можно воспроизвести с помощью деталей конструктора *Tinkertoy*®. Конечно, чтобы отразить всю сложность нашего разума, эта конструкция должна быть очень большой – но что, если все-таки можно создать такой сложный механизм, подобный компьютеру для игры в крестики-нолики из 10 тысяч деталей *Tinkertoy*? Теоретически наш разум мог бы материализоваться даже в движении листьев или в полете стаи птиц. В 1714 г. немецкий философ и математик Готфрид Лейбниц описал в трактате «Монадология» воображаемую машину размером с мельницу, способную мыслить и чувствовать. Если мы войдем внутрь, как в мельницу, то при осмотре «не найдем ничего внутри ее, кроме частей, толкающих одна другую, и никогда не найдем ничего такого, чем бы можно было объяснить восприятие»[4]. Аналогичным образом в будущем могут развиться сущности с ИИ, обладающие сознанием, даже не состоящие из органической материи.

Философ Ник Бостром (р. 1973) рассуждает об электронной клетке мозга: «Клетка мозга – это физический объект с определенными характеристиками. Если мы научимся воспроизводить их в электронном виде, тогда, несомненно, электронная мозговая клетка сможет выполнять те же функции, что и органическая. И почему бы полученной в итоге системе не обладать сознанием, подобно человеческому мозгу?»

Робототехник Ханс Моравек (р. 1948) пишет: «Мы сознательные существа, смоделированные на основе нейронного оборудования, но сознание проявляется лишь в интерпретации процессов, происходящих в сетях нейронов. Не реальные химические сигналы, а высокоуровневая интерпретация совокупности этих сигналов – вот что отличает сознание от прочих видов обработки данных, подобно представлению о ценности долларовой банкноты».

Возможно, ваш мозг точно так же мог бы функционировать, даже будучи разделенным на сотню маленьких коробочек, расположенных на большом расстоянии друг от друга и соединенных проводами или оптическими волокнами. Представьте себе, что левое и правое полушария вашего мозга находятся в километре друг от друга и соединены искусственным мозолистым телом. Вы ведь все равно останетесь собой, не так ли?

СМ. ТАКЖЕ Крестики-нолики (1300 до н. э.), «Левиафан» Гоббса (1651), Механическая утка де Вокансона (1738), Поиски души (1907), «Гигантские мозги, или Машины, которые думают» (1949), Жизнь в симуляции (1967)

[4] Пер. Е. Н. Боброва.

Если мы считаем, что сознание – результат комбинаций и динамических взаимосвязей нейронов и других клеток мозга, то наши мысли, эмоции и воспоминания могут быть воспроизведены в движениях веток и листьев или в полете стаи птиц. (Акварель Бенжависы Руангвари.)

Станок для написания книг в Лагадо. 1726

В «Путешествиях Гулливера», знаменитом романе англо-ирландского писателя Джонатана Свифта (1667–1745), изданном в 1726 г., упоминается механический станок для творческой работы. Возможно, это первое ИИ-устройство, подробно описанное в художественной литературе. Во время пребывания Гулливера в вымышленном городе Лагадо профессор показывает ему станок для создания художественных и технических книг и порождения интересных идей и объясняет, что «благодаря его изобретению самый невежественный человек с помощью умеренных затрат и небольших физических усилий может писать книги по философии, поэзии, политике, праву, математике и богословию при полном отсутствии эрудиции и таланта»[5].

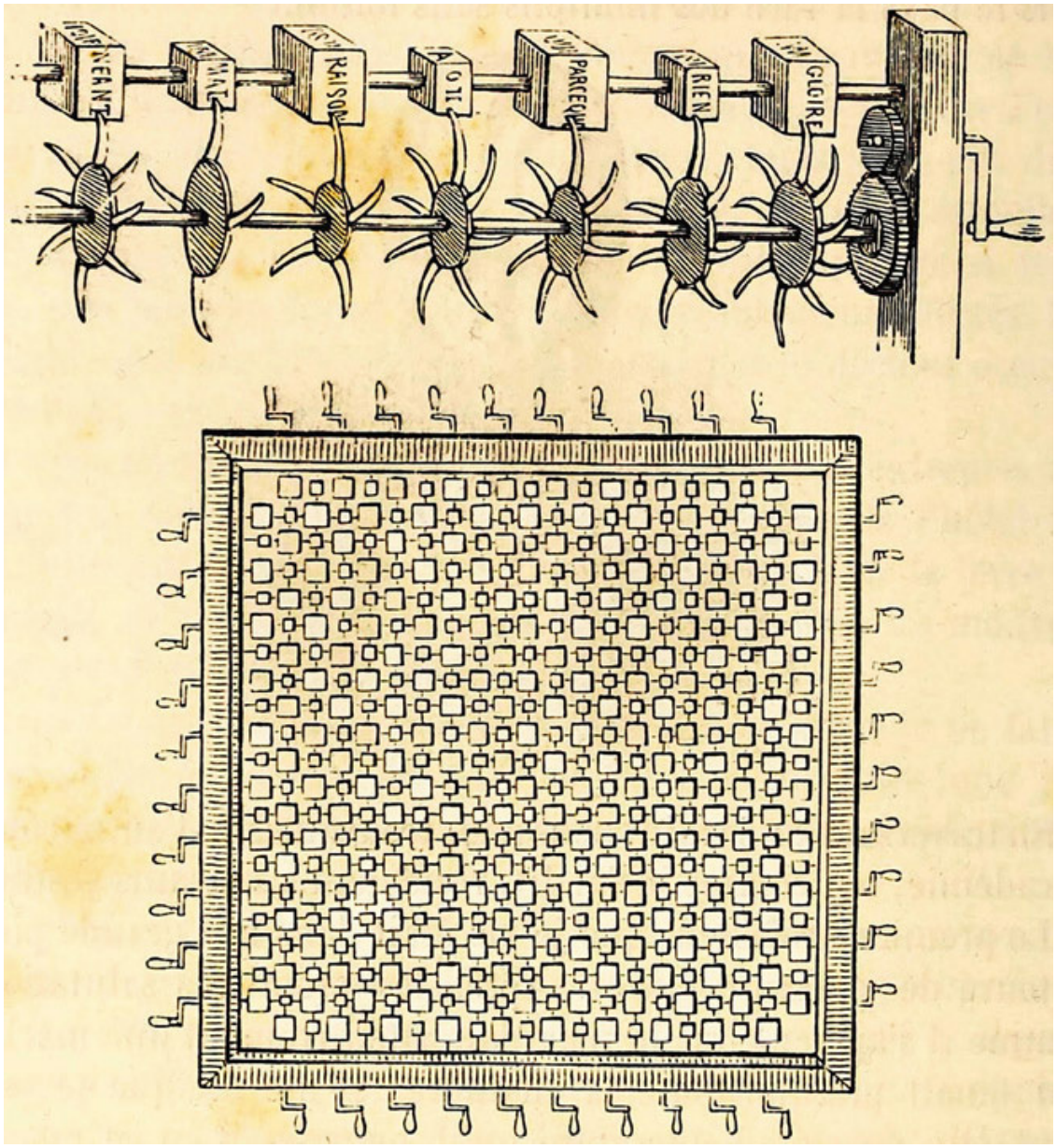
Профессор подводит Гулливера к раме, занимающей площадь в 20 квадратных футов, поверхность которой состоит из множества деревянных дощечек, «сцепленных между собой тонкими проволоками». Со всех сторон дощечки оклеены бумагой, на которой «написаны все слова их языка в различных наклонениях, временах и падежах, но без всякого порядка».

Гулливер так описывает работу устройства: «По его команде каждый ученик взялся за железную рукоятку, которые в числе сорока были вставлены по краям рамы, и быстро повернул ее, после чего расположение слов совершенно изменилось. Тогда профессор приказал тридцати шести ученикам медленно читать образовавшиеся строки в том порядке, в каком они разместились в раме; если случалось, что три или четыре слова составляли часть фразы, ее диктовали остальным четырем ученикам, исполнявшим роль писцов… После каждого оборота слова принимали все новое расположение, по мере того как квадратики переворачивались с одной стороны на другую».

Профессор Стэнфорда Эрик Вайс пишет: «Назначение [машины], заявления ее просвещенного изобретателя, его призывы к государственному финансированию, а также тот факт, что с ней работали ученики, позволяют отнести ее к ранним попыткам создания искусственного интеллекта и приводить в качестве примера из истории этой дисциплины». К более поздним примерам комбинаторного, случайного или смоделированного творчества в реальном мире относится *Racter* – компьютерная программа, которая сгенерировала тексты для книги «Борода полицейского наполовину сконструирована», опубликованной в 1984 г.

СМ. ТАКЖЕ *Ars Magna* Раймунда Луллия (ок. 1305), Машинное творчество (1821), «Кибернетическая проницательность» (1968)

[5] Пер. А. А. Франковского.



Станок для написания книг из Лагадо, изображенный французским художником Жаном Гранвилем (1803–1847). Иллюстрация из французского издания «Путешествий Гулливера» 1856 г.

Механическая утка де Вокансона. 1738

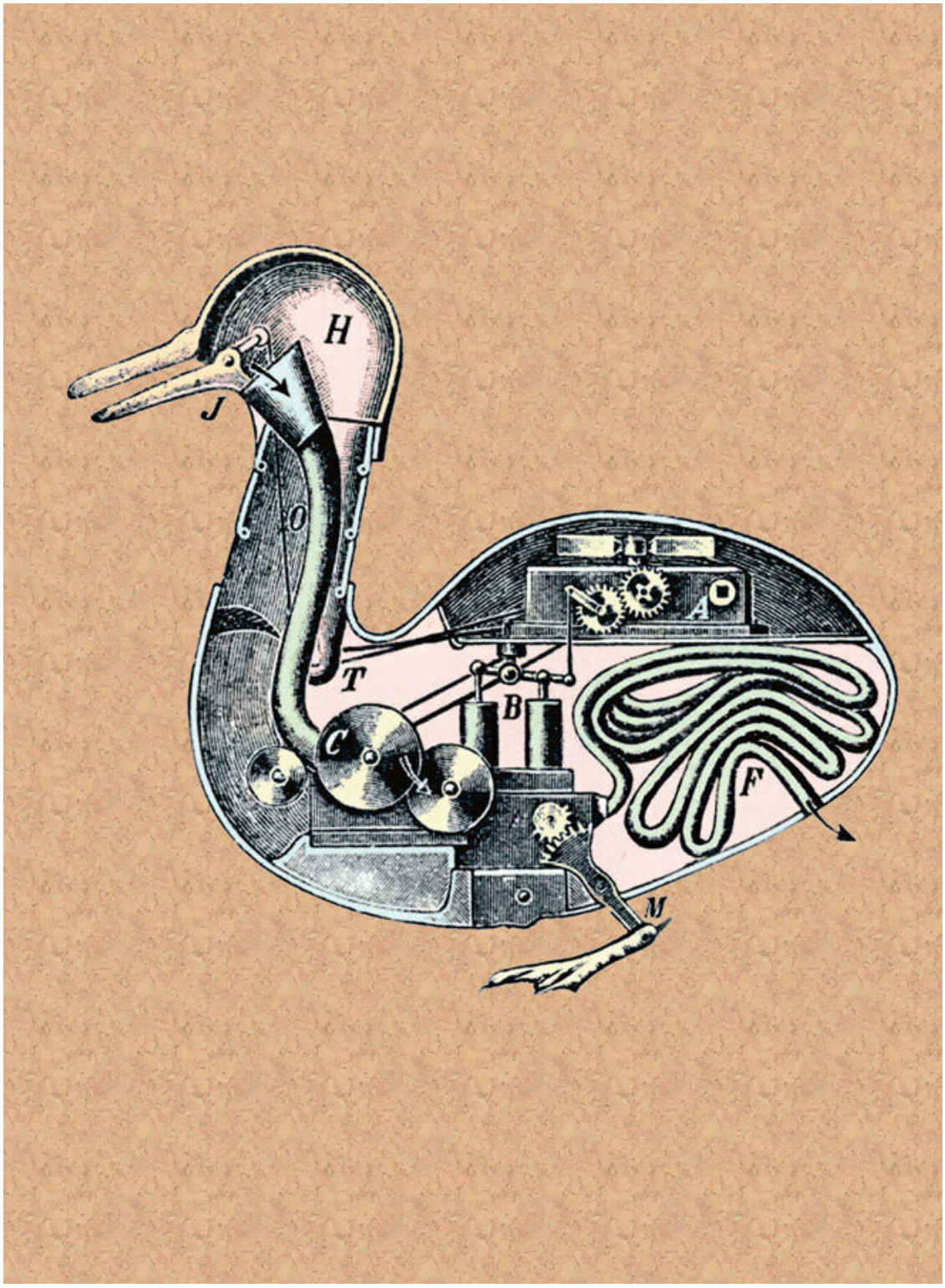
«В 1738 г. 29-летний французский часовщик Жак де Вокансон (1709–1782) выставил в саду Тюильри экспонат, который можно назвать одним из самых знаменитых роботов всех времен», – пишет американский нейробиолог Пол Глимчер. У утки де Вокансона были сотни движущихся частей и перьев. Она крутила головой, мутила воду клювом, хлопала крыльями, крякала, заглатывала пищу из рук смотрителя и весьма реалистично совершала множество других действий. Спустя несколько минут остатки «переваренной» пищи выводились наружу. Конечно, на самом деле утка ничего не переваривала: в ее хвост заранее незаметно загружалось некое подобие экскрементов. И все же этот разносторонний автоматон спровоцировал дискуссии о границе между живым и чисто механическим, а также о том, насколько эта граница может размываться по мере того, как роботизированные сущности становятся все более универсальными.

Со временем интерес к знаменитой «переваривающей утке» только возрастал, и это странное существо даже упоминается в известном романе Томаса Пинчона «Мейсон и Диксон», опубликованном в 1997 г. Там механическая утка обретает сознание и преследует французского шеф-повара, терроризируя его своим *bec de la mort* (клювом смерти).

Де Вокансон также создал необыкновенного робота-флейтиста, который издавал звуки благодаря нескольким мехам, соединенным с тремя духовыми каналами. Шестеренки и кулачковые механизмы приводили в действие рычаги, управлявшие пальцами, языком и губами музыканта. По словам историка Джессики Рискин, механический флейтист «был первым примером того, что в „Энциклопедии“ Дидро определялось как андроид, – то есть антропоморфной фигурой, выполняющей человеческие действия». В 1740-х гг. де Вокансон изобрел нечто более практичное – ткацкие станки для изготовления шелка. Увы, из-за этого ткачи взбунтовались и забросали его на улице камнями.

Глимчер отмечает: «Утка де Вокансона подняла перед обществом XVIII века давние вопросы, которые и сегодня не дают покоя нейробиологам: достаточно ли одних только механических взаимодействий, происходящих внутри каждого человека, чтобы производить те сложные модели поведения, которые все мы демонстрируем? Что именно делает нас людьми: сложность нашего поведения или особенности взаимодействующей материи, которая, по-видимому, его определяет?»

СМ. ТАКЖЕ Механический парк в Эдене (ок. 1300), Робот-рыцарь Леонардо да Винчи (ок. 1495), Мельница Лейбница (1714), «Паровой человек в прериях» (1868), «Большой черный страус Электрического Боба» (1893)



Популярное изображение утки-автоматона де Вокансона, появившееся в журнале *Scientific American* 21 января 1899 г. Хотя изображенный здесь механизм не вполне соответствует реальному внутреннему устройству утки, стрелка верно указывает на выходное отверстие.

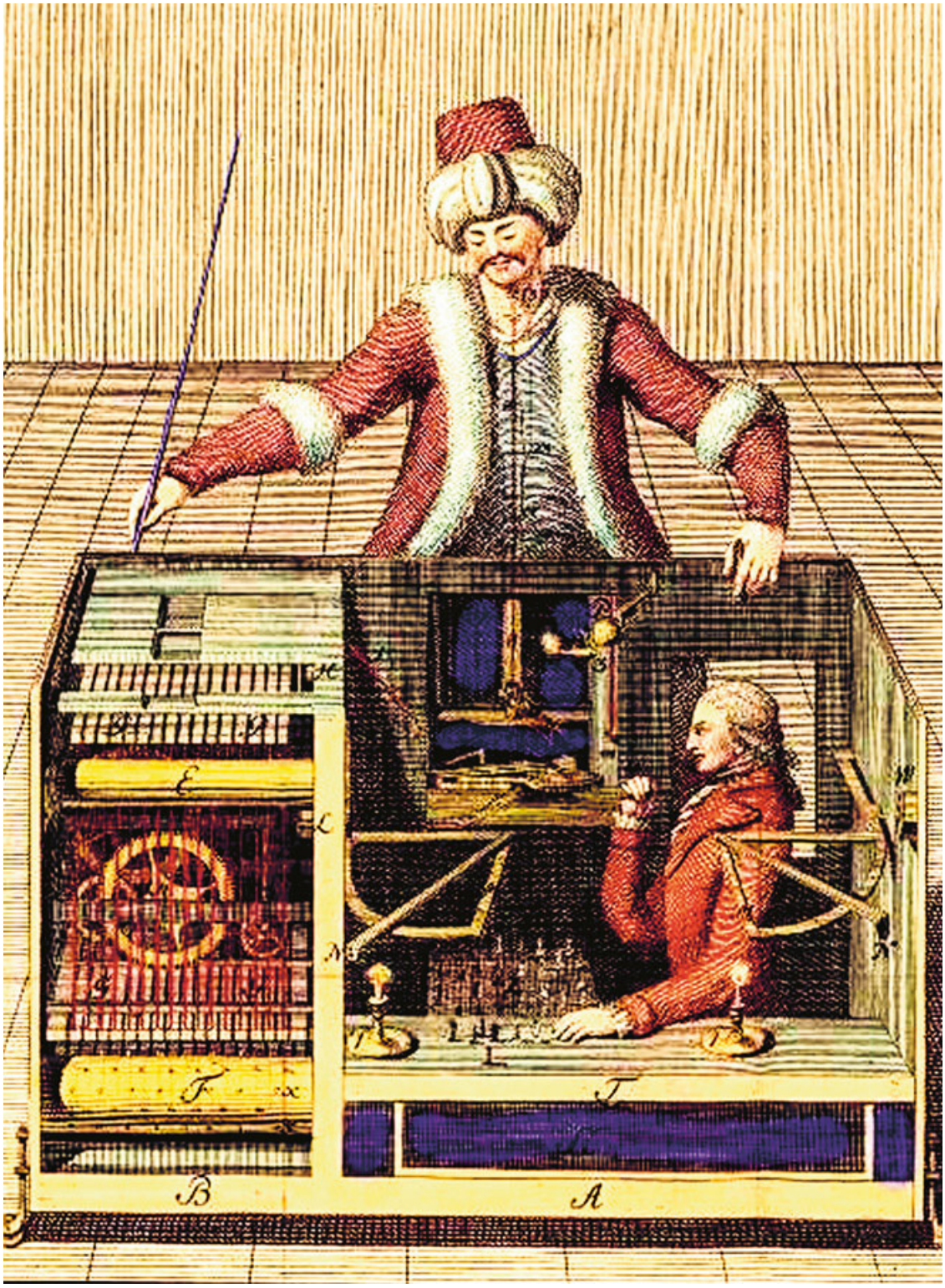
Механический турок. 1770

Механический турок представлял собой шахматиста-андроида. Его создал в 1770 г. венгерский изобретатель Вольфганг фон Кемпелен (1734–1804) в подарок императрице Марии-Терезии. Машина демонстрировала превосходную игру в шахматы и одерживала победы над самыми разными шахматистами в Европе и Америке, включая таких известных личностей, как Наполеон Бонапарт и Бенджамин Франклин. Андроид размером с обычного человека – с черной бородой, облаченный в мантию и тюрбан, – сидел за большим деревянным ящиком с шахматной доской и перемещал рукой фигуры. На протяжении многих лет секрет его действия не был известен широкой публике, но сегодня мы знаем, что в хитроумно устроенном ящике ловко скрывался живой человек – шахматист, который перемещал фигуры по доске с помощью магнитов и использовал рычаги для управления разными частями тела андроида. Чтобы добавить таинственности, перед игрой Кемпелен открывал дверцы ящика, демонстрируя внутренний заводной механизм, за которым, казалось, негде было спрятаться человеку. Хотя многие догадывались, что турок – это лишь хитрый фокус, автомат все же заставлял людей задуматься, на что еще способны машины и какой человеческий труд они могли бы взять на себя.

О работе механического шахматиста было написано множество статей с неверными догадками. Например, Эдгар Аллан По ошибочно предполагал, что человек сидел внутри самого андроида. Интересно, что один из отцов современных компьютеров Чарльз Бэббидж, вероятно, вдохновлялся турком Кемпелена: начав разрабатывать механические вычислительные машины, он задавался вопросом, могут ли машины «думать» или, по крайней мере, выполнять очень сложные вычисления.

Элла Мортон отмечает: «Хотя успех [турка], по сути, был обусловлен действиями человека и элементами старомодной магии, его убедительно механическая природа вызывала удивление и беспокойство. Появившись в самый разгар промышленной революции, турок поднял непростые вопросы о природе автоматизации и возможности создания мыслящих машин. Дело в том, что турок, работавший, как казалось, на заводных механизмах… опровергал идею о том, что шахматы – „территория одного лишь разума“».

СМ. ТАКЖЕ Механический компьютер Бэббиджа (1822), «Слоны не играют в шахматы» (1990), Шашки и искусственный интеллект (1994), *Deep Blue* обыгрывает чемпиона мира по шахматам (1997)



Внутренний механизм турка согласно гипотезе Йозефа фон Ракница (1744–1818). (Из книги *Über den schachspieler des herrn von Kempelen und dessen nachbildung* [О Шахматисте господина фон Кемпелена и его копии], Лейпциг и Дрезден, 1789.)

Автоматоны Жаке-Дро. 1774

«Это могли быть огромные марионетки или высокие куклы-манекены, брошенные в панике, как я догадался, из-за какой-то чумы… которая пронеслась по городу и заставила жителей покинуть его. Я остался наедине с этими симулякрами любви… завороженный неподвижными блестящими глазами всех этих автоматонов», – писал романист Жан Лоррен (1845–1906).

Эти жутковатые мысли о натуралистичных автоматах напоминают о давнем увлечении человечества роботоподобными существами и об особой группе таких автоматонов, созданных в XVIII в. Учитывая сложность и программируемость этих андроидов, их можно назвать далекими предками компьютеров. Речь идет о трех устройствах, созданных часовщиком Пьером Жаке-Дро (1721–1790) между 1768 и 1774 гг. и привлекавших толпы поклонников: мальчик-писарь (состоящий примерно из 6000 деталей), девушка-музыкант (2500 деталей) и мальчик-художник (2000 деталей). Мальчик-писарь, макая гусиное перо в чернила, писал тексты длиной до сорока символов; содержание текстов программировалось на особом зубчатом диске. Время от времени андроид опускал перо в чернильницу, а выводя буквы, следил глазами за написанным.

Механическая девушка-музыкант играла на миниатюрном органе, по-настоящему нажимая на клавиши пальцами. Будто оживая, она весьма естественно двигалась, поводила головой и следила взглядом за руками. Она «дышала» даже до и после своих выступлений, а во время игры ее тело словно выражало эмоции, соответствующие мелодии. Мальчик-художник умел делать четыре наброска: собаку, портрет Людовика XV, Купидона, управляющего колесницей, и королевскую чету.

Эти автоматоны примечательны тем, что их механизмы скрыты внутри самих фигур (а не, например, в соседнем предмете мебели), поэтому их программируемость, реалистичность и миниатюрные размеры особенно восхищают. Конструировать их Пьеру Жаке-Дро помогали его сын Анри-Луи и приемный сын Жан-Фредерик Лешо, талантливый механик. По некоторым данным, позднее изобретатель изготовил две руки-протеза для человека с врожденными патологиями. По слухам, эти руки, облаченные в белые перчатки, были настолько универсальны, что с их помощью можно было писать и рисовать.

СМ. ТАКЖЕ Автоматы Аль-Джазари (1206), Механический парк в Эдене (ок. 1300), Религиозные автоматы (1352), Механическая утка де Вокансона (1738)

Механический мальчик-писарь, созданный Жаке-Дро, в Музее искусства и истории в Невшателе (Швейцария). Автоматоны работают и по сей день.

«Франкенштейн». 1818

«“Франкенштейн” был написан во время первой промышленной революции, – пишет Паоло Галло, директор по кадрам Всемирного экономического форума. – Это был период грандиозных перемен, которые у многих вызывали беспокойство и замешательство. В романе поднимались важные вопросы о взаимоотношениях человека и технологий: не создаем ли мы монстра, которого не сможем контролировать? Не утрачиваем ли мы человечность, сострадание, способность сопереживать и испытывать эмоции?»

Опасность, которую представляет особый вид ИИ, – одна из главных тем романа Мэри Шелли (1797–1851) «Франкенштейн, или Современный Прометей» (1818). Главный герой, ученый Виктор Франкенштейн, похищает со скотобоен и кладбищ куски неживой плоти, чтобы создать из них существо, которое впоследствии оживляет «искрой жизни». Он смотрит на свое творение как на эксперимент по достижению бессмертия: «Мне первому предстояло преодолеть грань жизни и смерти и озарить наш темный мир ослепительным светом. Новая порода людей благословит меня как своего создателя… Раз я научился оживлять мертвую материю, рассуждал я, со временем… я сумею также давать вторую жизнь телу, которое смерть уже обрекла на исчезновение»[6].

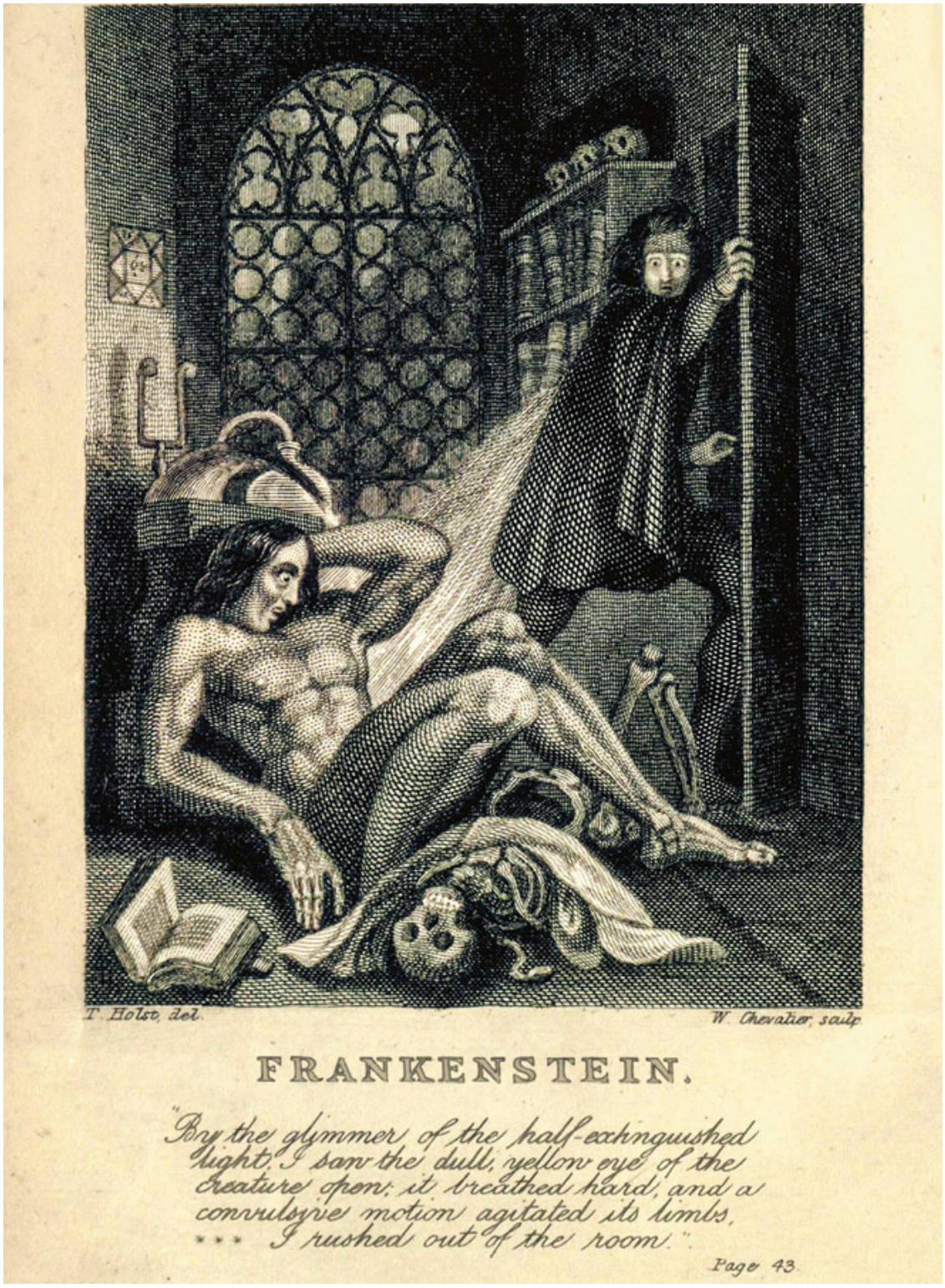
Мэри Шелли закончила свой роман, когда ей было 19 лет. Сюжетная идея пришла к писательнице во сне. В то время в Европе были популярны теории, согласно которым электричество играет большую роль в биологии и может быть использовано для оживления мертвых тканей. Известно также, что примерно в 1803 г. в Лондоне итальянский физик Джованни Альдини (1762–1834) предпринимал многочисленные публичные попытки оживления людей с помощью электричества.

Мотивами смерти и разрушения пронизан весь роман, в котором погибают многие персонажи. Так, Виктор уничтожает незавершенную женщину – подругу для монстра, а сам монстр (которого, кстати, в романе ни разу не называют Франкенштейном) убивает Элизабет, жену ученого. Виктор преследует свое творение до Северного полюса, где сам он умирает, а монстр клянется сжечь себя на своем погребальном костре.

Журналист Даниэль Д’Аддарио отмечает: «“Франкенштейн” строится на идее, что люди склонны отвергать искусственный разум как странный и неестественный. Во многом это объясняется отвратительным видом чудовища Франкенштейна… Но что произойдет, когда ИИ явится нам в более привлекательной и практичной сборке?»

СМ. ТАКЖЕ Талос (ок. 400 до н. э.), Голем (1580), «Паровой человек в прериях» (1868), *R.U.R.* (1920)

[6] Пер. З. Е. Александровой.



Фронтиспис издания «Франкенштейна» 1831 г., выпущенного в Лондоне издательством Генри Колберна и Ричарда Бентли.

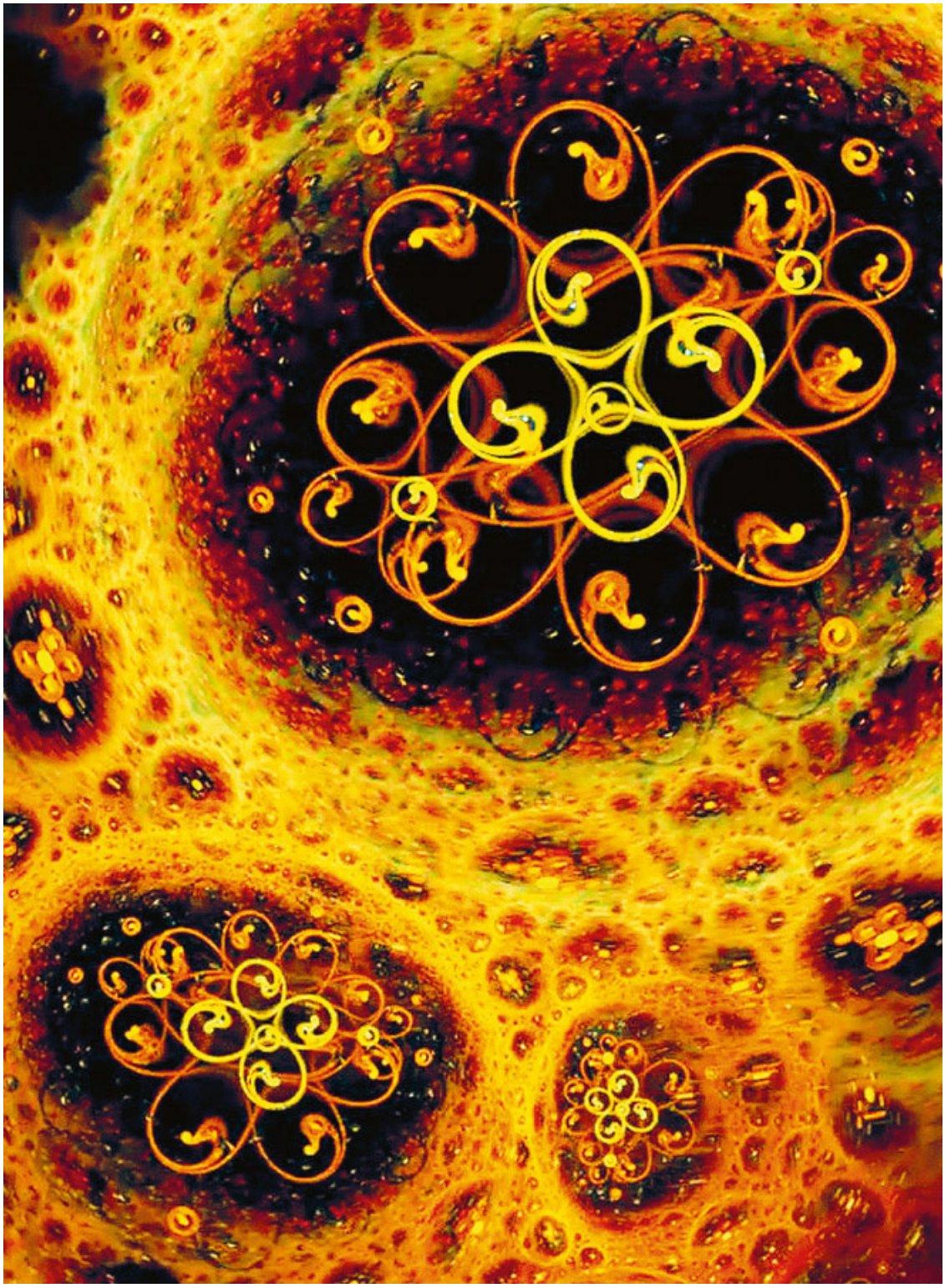
Машинное творчество. 1821

«Наше общество ревниво оберегает все, что связано с творчеством, – пишут Саймон Колтон и Герейнт Уиггинс, исследователи из вычислительного подразделения Голдсмитского колледжа. – Человеческое творчество опирается на весь спектр наших умственных способностей, поэтому его моделирование – серьезная техническая проблема. Нам кажется справедливым считать машинное творчество пределом возможностей ИИ – вероятно, его последним пределом».

У термина «машинное творчество» есть несколько значений. Мы подразумеваем под ним область разработки ИИ, в которой творчество моделируется с помощью компьютера или другой машины. Результаты такого моделирования часто кажутся оригинальными и потенциально полезными. К машинному творчеству также относят программы, расширяющие творческие возможности человека. Например, с помощью искусственных нейросетей и других инструментов можно писать музыку или картины в стиле известных художников. Порождающие состязательные сети (GAN, комбинации из двух соперничающих нейросетей) создают фотореалистичные изображения смоделированных лиц, цветов, птиц и интерьеров. С помощью машинного творчества создают новые формы визуального искусства, поэтические и прозаические произведения, математические теоремы, кулинарные рецепты, объекты патентования, игры и шахматные задачи, конструкции антенн и теплообменников и многое другое. Компьютер может порождать такие объекты, которые, будь они созданы человеком, считались бы актом творчества.

К числу ранних и простых примеров машинного творчества относится *компониум* – механический музыкальный орган, способный создавать огромное число вариаций на музыкальную тему. Его сконструировал в 1821 г. изобретатель Дитрих Винкель (1777–1826). У компониума было два цилиндра, которые поочередно исполняли два такта из случайно выбранного музыкального произведения. Выбор вариаций определял специальный маховик, выполнявший функцию программирующего устройства. По словам Винкеля, если бы каждое выступление длилось пять минут, то компониуму потребовалось бы больше 138 триллионов лет, чтобы воспроизвести все возможные музыкальные комбинации!

СМ. ТАКЖЕ *Ars Magna* Раймунда Луллия (ок. 1305), Станок для написания книг в Лагадо (1726), «Кибернетическая проницательность» (1968), Генетические алгоритмы (1975), Компьютерное искусство и *DeepDream* (2015)



«Электрическая овца». Эта абстрактная заставка для экрана была сгенерирована в рамках коллективного художественного проекта *Electric Sheep*, основанного Скоттом Дрейвсом. Так называемые «овцы» (узоры), наиболее популярные среди пользователей, живут дольше и размножаются при помощи генетического алгоритма, скрещиваясь и мутируя.

Механический компьютер Бэббиджа. 1822

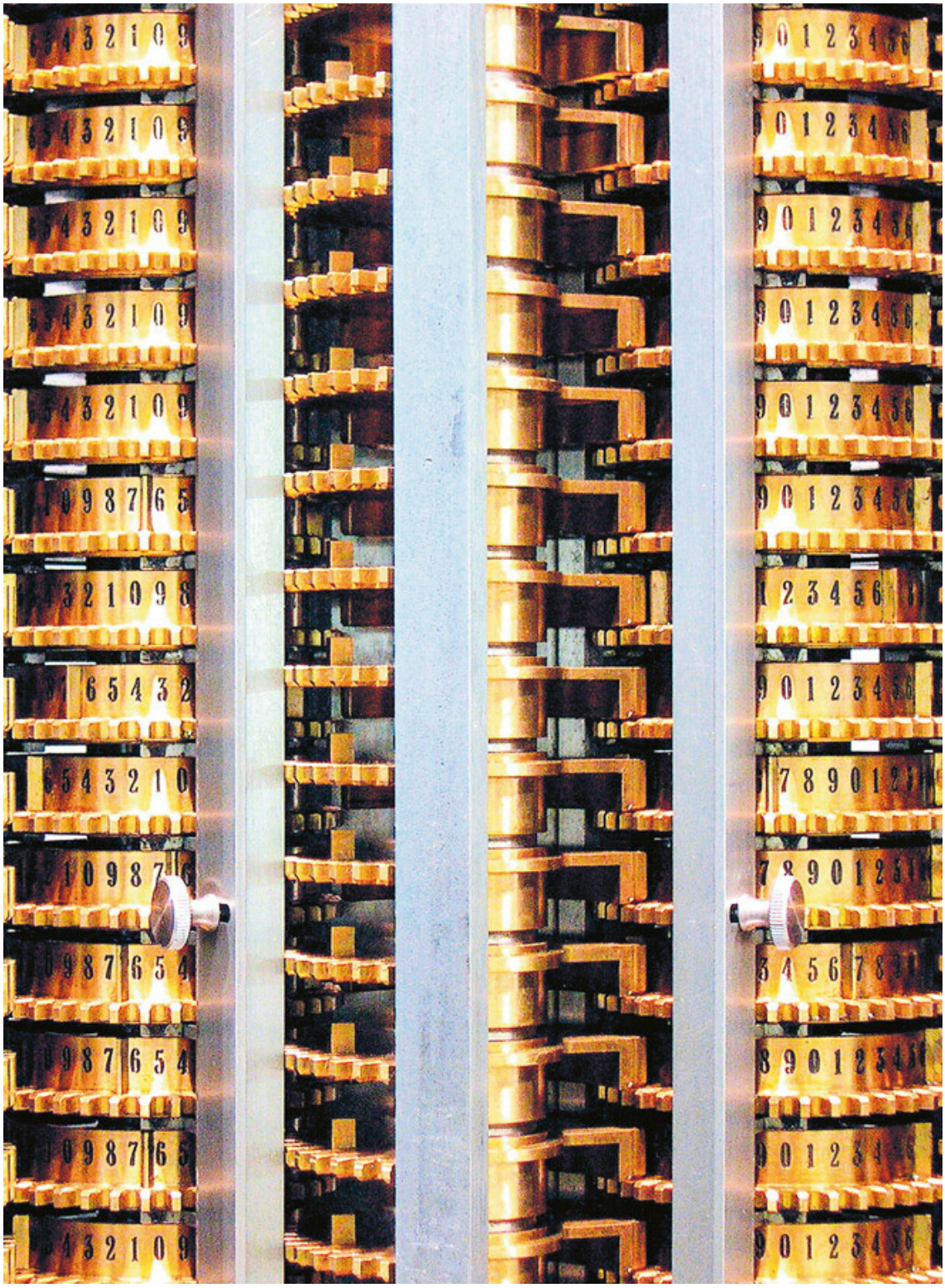
Чарльз Бэббидж (1791–1871) – английский математик, статистик и изобретатель. В 1819 г. он увидел механического турка, когда тот гастролировал по Англии и обыгрывал живых людей в шахматы. Конечно, Бэббидж наверняка догадывался, что турок был каким-то фокусом, но многие считают, что этот андроид вдохновил математика на размышления о других, более практичных думающих машинах и это стало первым шагом на пути к ИИ.

Бэббиджа часто называют самым выдающимся математиком-инженером в предыстории компьютеров. В частности, он создал грандиозный механический калькулятор с ручным приводом – прообраз современного компьютера. Бэббидж считал, что этот прибор будет особенно полезен при составлении математических таблиц, но его беспокоили ошибки, которые могут допустить люди, переписывая результаты с металлических барабанов (всего их было 31). Сейчас мы понимаем, что Бэббидж примерно на столетие опередил свое время и что общество и технологии его эпохи были недостаточно развиты для воплощения высоких мечтаний ученого.

Разностная машина Бэббиджа, работа над которой началась в 1822 г., но так и не была завершена, предназначалась для вычисления значений многочленных функций и состояла примерно из 25 тысяч механических деталей. Ученый также планировал создать устройство более общего назначения – аналитическую машину, которую можно было бы программировать с помощью перфокарт и у которой имелись бы отдельные участки для вычисления и сохранения результатов. Согласно расчетам, аналитическая машина, способная хранить тысячу пятидесятизначных чисел, имела бы длину более тридцати метров. Ада Лавлейс (1815–1852), дочь английского поэта Байрона, составила подробное описание аналитической машины и написала для нее программу. Хотя Аде помогал в этом Бэббидж, многие считают ее первым программистом в истории.

В 1990 г. Уильям Гибсон и Брюс Стерлинг написали роман «Машина различий», в котором изобразили возможные последствия появления механических устройств Бэббиджа в викторианском обществе. В конце романа описывается альтернативный 1991 г., в котором осознающий себя компьютер эволюционировал и, по-видимому, стал рассказчиком книги.

СМ. ТАКЖЕ Счеты (ок. 190 до н. э.), Механический турок (1770), ЭНИАК (1946)



Рабочая модель части разностной машины Чарльза Бэббиджа. В настоящее время хранится в лондонском Музее науки.

«Мастер красоты». 1844

«Мастер красоты» Натаниэля Готорна (1804–1864) – это первый в мире рассказ о роботе-насекомом, завораживающе прекрасный и примечательный поднятыми вопросами об ИИ и отношении к нему человечества. Рассказ, опубликованный в 1844 г. – задолго до изобретения электрической лампочки, – повествует о жизни талантливого мастера Оуэна Уорленда, который работает в часовой мастерской. Оуэн – тонко чувствующий молодой человек, тайно влюбленный в Энни Ховенден, дочь старого часовщика, хозяина мастерской. С самого детства ему хотелось «воссоздать ту первозданную красоту, которая присуща полету птиц и движениям маленьких животных»[7].

В конце концов Оуэну удается сделать механическую бабочку. Старый часовщик обнаруживает первую модель и едва не разрушает механизм, «хрупкий и миниатюрный, как внутренности бабочки». Он резко заявляет: «Оуэн, в этих звеньях, винтиках и колесиках какая-то черная магия».

В заключительной сцене рассказа Оуэн решает показать Энни свое новое творение. Бабочка «уселась на кончик Эннииного пальца, трепеща дивными пурпурными в золотых крапинках крыльями, словно собираясь продолжить полет. Никакими словами не передать ее красоты, сотворенной из блеска, великолепия, нежно-переливчатой яркости. В ней был воплощен идеал живой бабочки, созданной самой Природой, – но не блеклой, что летает среди земных цветов, а той, что носится над райскими лугами, радуя маленьких ангелов и души умерших младенцев.

– Красавица! Красавица! – воскликнула Энни. – Она живая? Живая?»

Насекомое летает по комнате и порхает вокруг головы Энни. История заканчивается печально для бабочки: беззаботный младенец давит ее, превращая в кучку крошечных блестящих обломков. Однако у Оуэна наступает некое прозрение: он понимает, что красота бабочки вечна.

Любопытно, что в 2015 г. в США был выдан патент № US9046884 на роботизированную бабочку, способную чувствовать, реагировать и двигаться в соответствии с эмоциональным состоянием человека, чтобы улучшать его настроение.

СМ. ТАКЖЕ Механическая утка де Вокансона (1738), Автоматоны Жаке-Дро (1774), Тик-Ток (1907), «Искусственный разум» Спилберга (2001)

[7] Пер. Э. Л. Линецкой.

В рассказе «Мастер красоты» описано создание прекрасной и хрупкой механической бабочки, обладающей как реалистичными, так и мистическими свойствами.

Булева алгебра. 1854

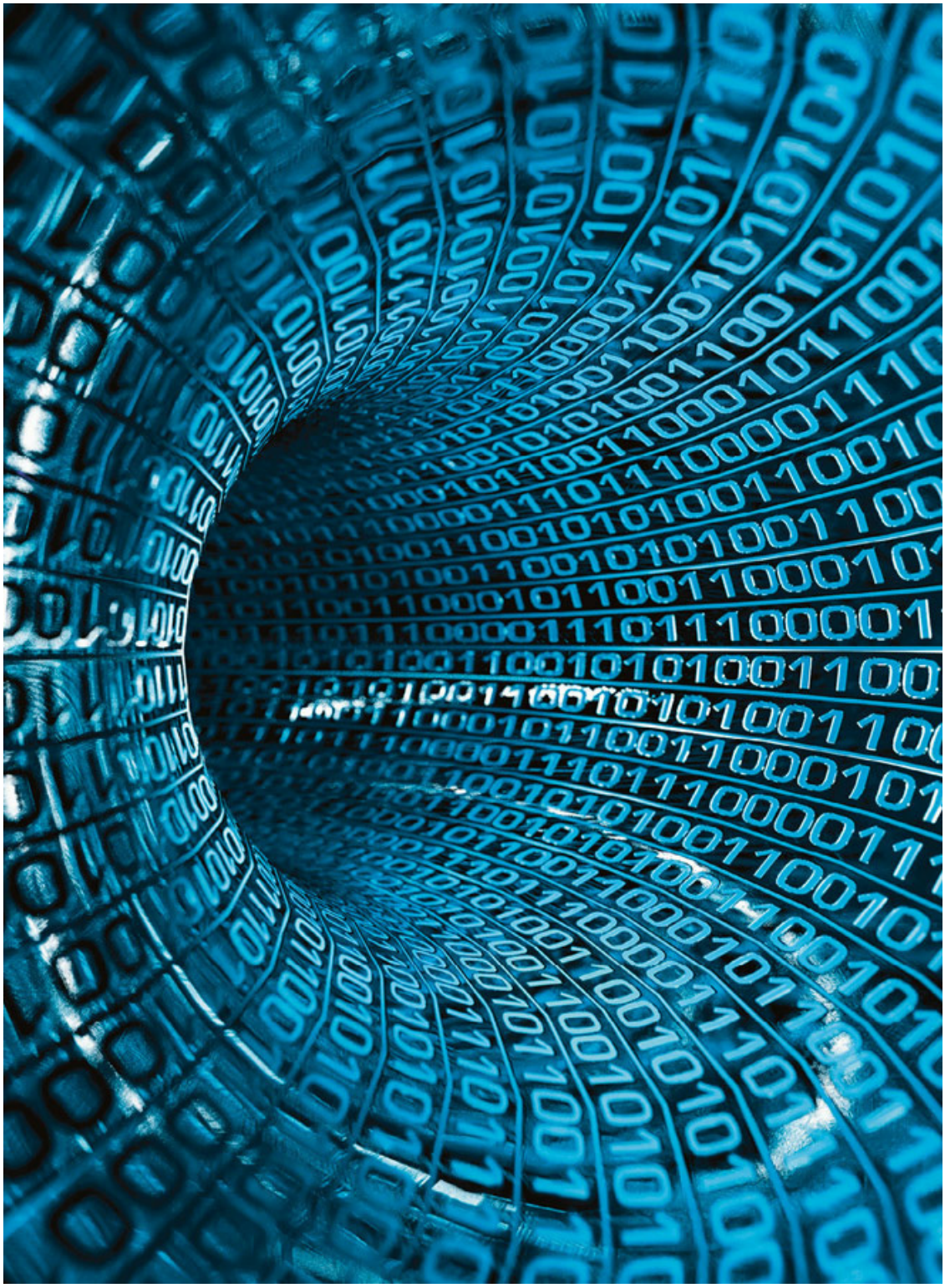
В 200-ю годовщину со дня рождения английского математика Джорджа Буля (1815–1864) журналист Джеймс Титкомб назвал его «первым теоретиком искусственного интеллекта, полагавшим, что человеческую мысль можно полностью свести к ряду математических правил и что машины способны заменить монотонный человеческий труд».

В своем главном труде Буль писал, что его целью было «изучить фундаментальные законы мыслительных операций, по которым выполняется рассуждение… собрать… некоторые возможные указания на природу и устройство человеческого разума». Его знаменитый трактат 1854 г. назывался «Исследование законов мышления, на которых основываются математические теории логики и вероятностей». Буля интересовала возможность свести логику к простой алгебре, используя только две величины, 0 и 1, и три простые операции: *и, или* и *не*. В наши дни булева алгебра находит широкое применение в телефонной коммутации и проектировании компьютеров.

Буль также указывал и на иную цель своих изысканий: «раскрыть тайные законы и взаимосвязи тех высоких функций мышления, с помощью которых мы постигаем и обдумываем все то, что находится за пределами чувственного познания мира и нас самих». Другой британский математик, Огастес де Морган (1806–1871), который ввел термин *математическая индукция*, высоко оценил работу Буля в своей посмертно опубликованной книге «Бюджет парадоксов»: «Система логики Буля – это лишь одно из многих свидетельств гениальности и терпения вместе взятых… В то, что символьные алгебраические операции, изобретенные для арифметических вычислений, подходят для выражения любых мыслительных действий и представляют собой грамматику и словарь всеохватной системы логики, никто не поверил бы, пока это не было доказано».

Примерно через семьдесят лет после смерти Буля американский математик Клод Шеннон (1916–2001), узнавший о булевой алгебре еще в студенческие годы, показал, как можно использовать ее для оптимизации систем телефонных коммутаторов. Он также продемонстрировал, что схемы с реле могут решать задачи булевой алгебры. Таким образом, Буль и Шеннон заложили одну из основ нашего цифрового века.

СМ. ТАКЖЕ «Органон» Аристотеля (ок. 350 до н. э.), Счеты (ок. 190 до н. э.), Нечеткая логика (1965)



Разрабатывая свою алгебру, Джордж Буль писал, что одной из его целей было «изучить фундаментальные законы мыслительных операций, по которым выполняется рассуждение».

«Дарвин среди машин». 1863

Английский писатель и энциклопедист Сэмюэл Батлер (1835–1902) одним из первых обрисовал ИИ-системы, которые могут появиться в будущем. В частности, он предвосхитил концепцию самосовершенствующегося машинного сверхразума и предсказал его потенциальные риски. В своем поразительном эссе 1863 г. «Дарвин среди машин» Батлер рассуждает о будущем «механической жизни»: «Мы сами создаем себе преемников; мы день ото дня делаем их физическую организацию все красивее и утонченнее; мы день за днем наделяем их все большей властью и снабжаем всевозможными хитроумными инструментами, формируя ту саморегулируемую, автономную силу, которая станет для них тем же, чем интеллект был для человеческой расы. Через несколько веков мы начнем считать себя низшей расой».

С поразительной проницательностью Батлер рисует картину постепенного подавления машинами человеческой воли: «Мы все больше от них зависим; все больше людей каждодневно становятся рабами, чтобы о них заботиться… посвящая все свои жизненные силы развитию механической жизни… Придет время, когда машины обретут реальную власть над миром и его обитателями…»

В «Книге машин» (1872) Батлер размышляет о том, что моллюск едва ли обладает особым сознанием, однако у человека сознание появилось в процессе эволюции. Точно так же, по мнению автора, машины смогут развить у себя сознание, а потому он призывает нас «поразмыслить над необычайным прогрессом, которого машины достигли за последнюю пару сотен лет, и обратить внимание на то, как медленно эволюционируют животные и растения. Более высокоорганизованные машины – это творения даже не вчерашнего дня, а последних пяти минут…».

Идеи Батлера эхом отозвались в XX в., когда отец кибернетики Норберт Винер (1894–1964) написал: «Если мы движемся к созданию машин, которые учатся и оттачивают свое поведение посредством опыта, мы должны признать: каждый новый уровень независимости, который мы открываем машине, становится уровнем возможного неповиновения нашим требованиям. Джинн не захочет добровольно вернуться в свою бутылку, и у нас нет никаких оснований ожидать, что машины будут к нам благосклонны».

Сегодня, когда технологии XXI в. проникли во все сферы человеческой жизни, размышления Батлера и Винера об ИИ выглядят особенно пророческими.

СМ. ТАКЖЕ «Левиафан» Гоббса (1651), «Человеческое использование человеческих существ» (1950), Интеллектуальный взрыв (1965), Герметичный ящик для искусственного интеллекта (1993), Максимизатор скрепок (2003)

В книге «Дарвин среди машин» Сэмюэл Батлер писал: «Мы сами создаем себе преемников… Через несколько веков мы начнем считать себя низшей расой».

«Паровой человек в прериях». 1868

Одно из первых описаний механического человека в американских «десятицентовых романах» (недорогих книгах в мягкой обложке) можно найти в «Паровом человеке в прериях». Этот роман был написан уроженцем Огайо Эдвардом Эллисом (1840–1916) и многократно переиздавался с 1868 по 1904 г. Его главный герой, подросток-изобретатель по имени Джонни Брейнерд, создает трехметрового робота и возит его по Среднему Западу США. Паровой человек носит шляпу-дымоход, таскает повозку, ходит и бегает на шипованных ногах со скоростью до 60 миль в час. Джонни с друзьями и роботом путешествуют по прериям, преследуя бизонов, пугая индейцев и помогая добывать золото.

Эллис так описывает парового человека: «невероятно крупный, распухший, будто олдермен, – что в общем-то выглядело гармонично, учитывая его громадный рост». Роботу сделали широкий корпус, чтобы в нем помещались все механизмы. Кроме того, «лицо было изготовлено из железа… с парой жутких глаз и здоровенным ухмыляющимся ртом… Походка казалась естественной, но во время бега осанка отличалась от человеческой».

Образ парового человека был вдохновлен реальным паровым человекоподобным роботом, запатентованным в 1868 г. американскими изобретателями Цадоком Дедериком и Айзеком Грассом. Увлекательные романы Эллиса представляют собой ранние образцы жанра «эдисонады»: в них обычно фигурирует молодой изобретатель, который спасается от опасностей благодаря своей смекалке. Историк Эндрю Липтак отмечает: «Рассказывая о приключениях на малоосвоенных землях, Эллис описывал повседневность людей, живущих на краю неизвестности. Это напоминает современные сюжеты, разворачивающиеся на далеких рубежах Солнечной системы… Книги Эллиса позволяют погрузиться в мировоззрение той эпохи и понять, каким тогда виделось будущее и насколько с тех пор изменился мир».

СМ. ТАКЖЕ «Большой черный страус Электрического Боба» (1893), Тик-Ток (1907), Подвижный робот Электро (1939)

Обложка книги «Паровой человек в прериях» Эдварда Эллиса (1868).

Ханойская башня. 1883

Игра «Ханойская башня», изобретенная в 1883 г. французским математиком Франсуа Эдуаром Люка (1842–1891), с момента появления и по сей день вызывает любопытство во всем мире. Эта математическая головоломка состоит из нескольких дисков разного диаметра, которые нанизываются на любой из трех стержней. В исходной позиции диски нанизаны на один стержень в порядке уменьшения размера (самый маленький располагается наверху). В процессе игры можно переносить на другой стержень по одному диску за раз: брать верхний диск из любой стопки и помещать его на вершину любой другой стопки. Диск большего диаметра нельзя помещать поверх меньшего. Цель – переместить всю исходную стопку (обычно состоящую из семи или восьми дисков) на другой стержень. Минимальное число ходов составляет $2^n - 1$, где n – количество дисков.

Считается, что создателя игры вдохновила легенда об индийском храме, в котором жрецы-брамины постоянно переносят 64 золотых диска с одного алмазного стержня на другой, руководствуясь описанными выше правилами. Согласно этой легенде, когда будет сделан последний ход, наступит конец света. Заметим, что если бы жрецы могли перемещать по одному диску в секунду, то на $2^{64} - 1$, или 18 446 744 073 709 551 615 ходов потребовалось бы примерно 585 миллиардов лет – это примерно в 42 раза больше нынешнего расчетного возраста нашей Вселенной.

Головоломка «Ханойская башня» и ее многочисленные вариации используются в различных робототехнических тестах, поскольку позволяют легко оценить по унифицированной шкале интеграцию высокоуровневой логики с восприятием и действиями робота. В таких тестах ключевую роль играет планирование задач и движений (с участием одной или нескольких рук робота).

Для вариаций с тремя стержнями существуют простые пошаговые решения, поэтому игра часто используется на уроках программирования для обучения рекурсивным алгоритмам. Однако найти оптимальное решение задачи «Ханойская башня» (и ее вариаций) с большим количеством стержней не так просто. А для многоруких роботов приходится еще и рассчитывать траектории, позволяющие избежать столкновений.

СМ. ТАКЖЕ Крестики-нолики (ок. 1300 до н. э.), Механический турок (1770), Четыре в ряд (1988), Роботы собирают кубик Рубика (2018)

В игре «Ханойская башня» можно переносить диски на другие стержни по одному: брать верхний диск из какой-либо стопки и помещать его на вершину любой другой стопки. Нельзя класть больший диск поверх меньшего.

«Большой черный страус Электрического Боба». 1893

Серия романов Роберта Тумбса об Электрическом Бобе, подобно «Паровому человеку в прериях» Эдварда Эллиса, показывает, как в США конца XIX в. возрастал интерес к механическим устройствам в виде людей и животных. Главный герой книги «Большой черный страус Электрического Боба» (1893) – десятилетний технический гений по прозвищу Электрический Боб, потомок изобретателя телеграфа Сэмюэла Морзе (1791–1872). Боб живет неподалеку от Нью-Йорка и конструирует различные механизмы: уже упомянутого страуса, гигантского белого аллигатора и других животных, которых можно использовать как транспорт. Как правило, эти роботы снабжены боеприпасами, имеют броню и умеют передвигаться по пересеченной местности.

Главному герою приходит в голову, что большой электрический страус мог бы перенести его с друзьями через скалистую пустыню на юго-западе США, позволив им избежать встречи со змеями. Боб тщательно изучает анатомию и физиологию страусовых и конструирует идеальное средство передвижения: «Большая голова страуса возвышалась почти на тридцать футов над землей. Центр тела находился в двадцати футах от земли, шея имела в длину около восьми футов… Энергия поступала от мощных аккумуляторных батарей, расположенных в теле птицы аккурат между бедрами. Они обеспечивали скорость от двадцати до сорока миль в час – в зависимости от свойств почвы, по которой мы передвигались».

Примечательно, что автор делится с читателем множеством технических подробностей. Например, в романе даются описания всех материалов и деталей механической птицы, включая полые стальные ноги и пуленепробиваемые алюминиевые крылья и хвост. Боб объясняет: «Здесь находятся канистры с водой, места для хранения провизии, боеприпасов и всего такого, а вот и наш пулемет… [который] состоит из увеличенного вращающегося барабана с двадцатью пятью винтовочными патронами и короткого тяжелого ствола. Чтобы стрелять, нужно поворачивать эту рукоятку».

Хотя подобная беллетристика не поднимает философских вопросов, связанных с созданием механических форм жизни, «эдисонады» показывают нам мировоззрение той эпохи – а также ее предрассудки, надежды и устремления.

СМ. ТАКЖЕ Механическая утка де Вокансона (1738), «Паровой человек в прериях» (1868), Тик-Ток (1907)

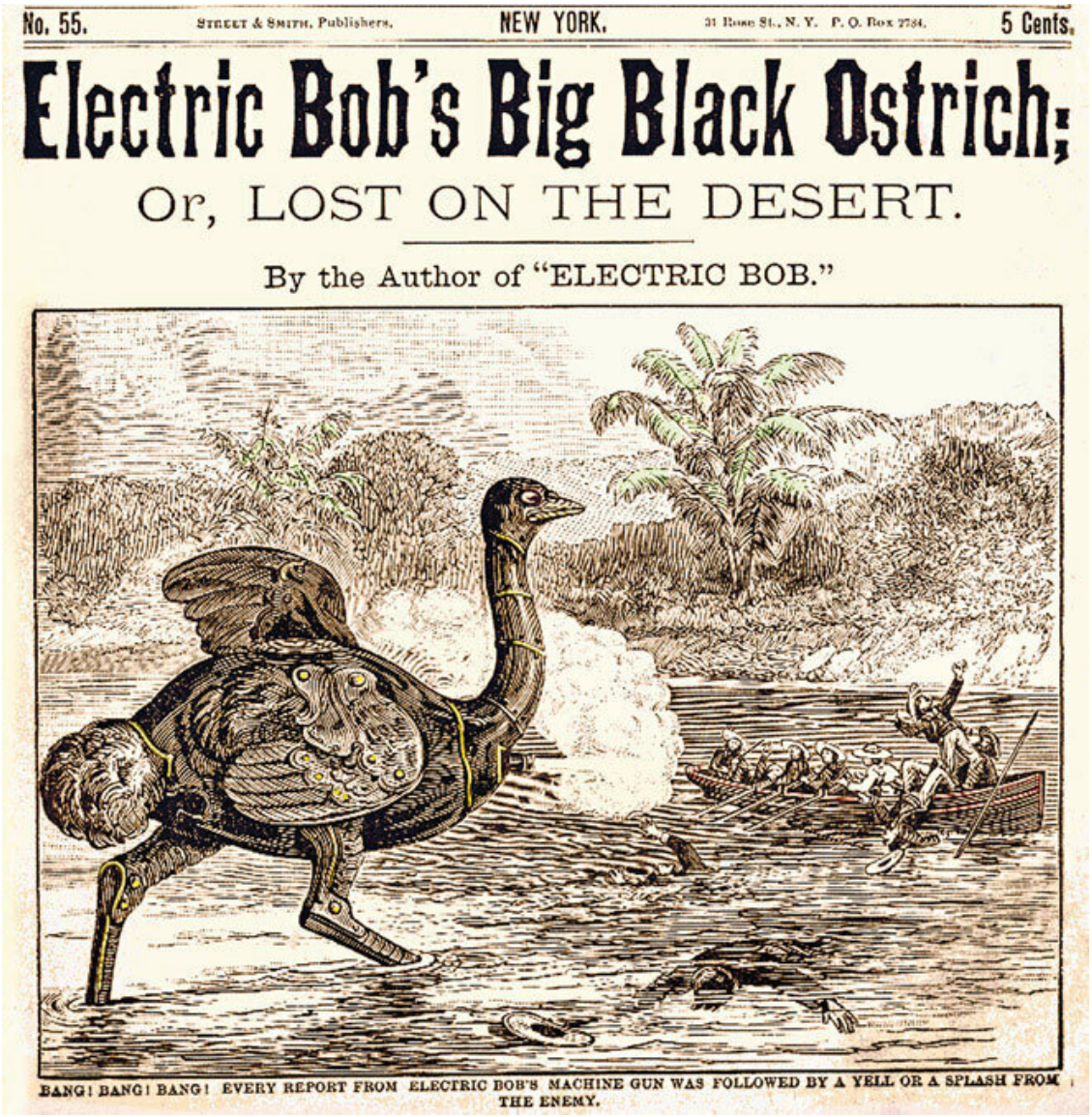


Иллюстрация к роману Роберта Тумбса «Большой черный страус Электрического Боба» из издательской серии New York Five Cent Library, 1893 г.

«Заимствованный разум» Теслы. 1898

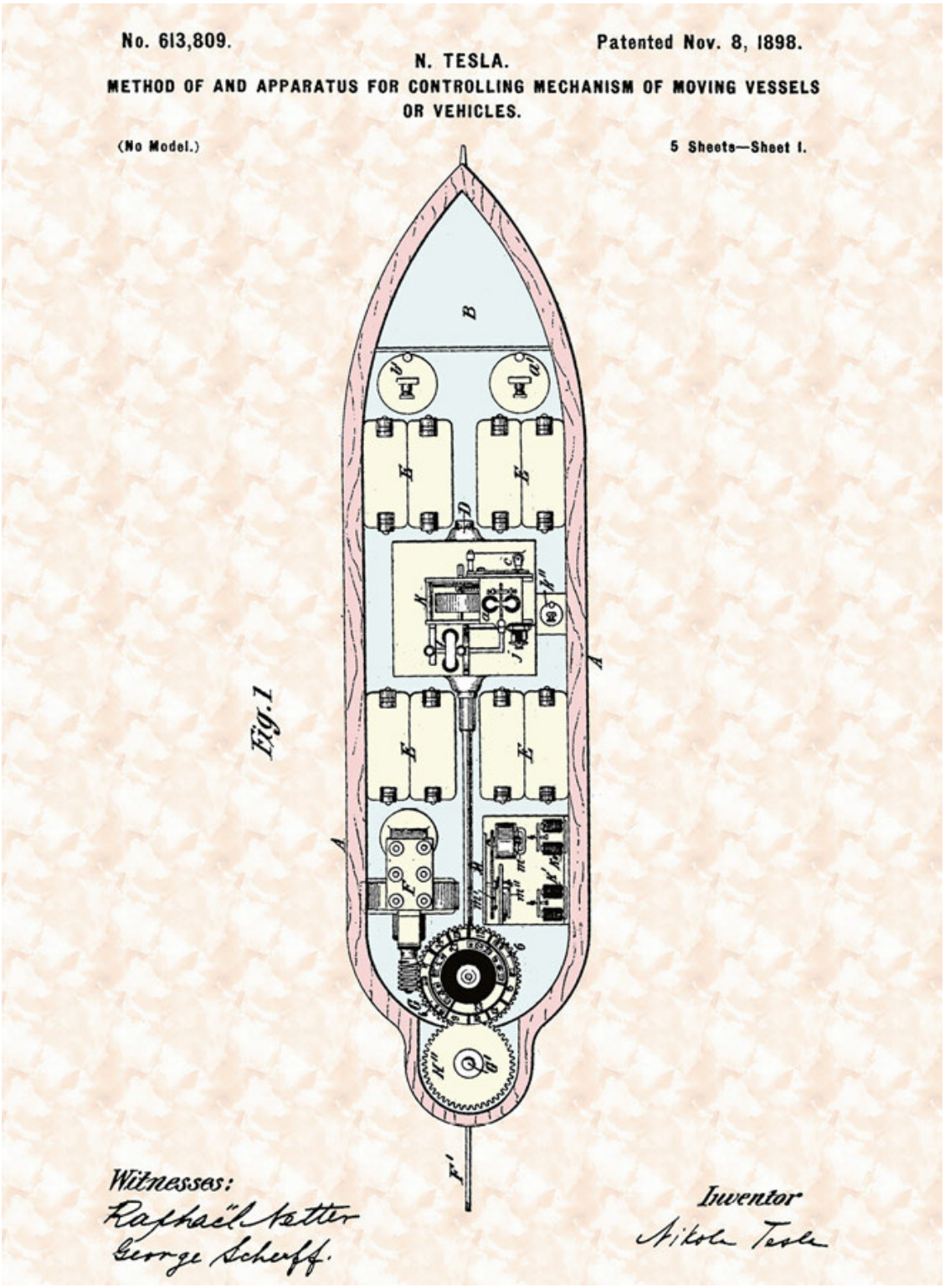
В 1898 г. сербско-американский изобретатель Никола Тесла (1856–1943) представил публике радиоуправляемую лодку, которая совершала маневры на глазах у изумленных зрителей. Некоторые даже думали, что в дело замешана магия, телепатия или дрессированная обезьяна. Когда репортер *New York Times* узнал о первом в мире радиоуправляемом судне, он предположил, что изобретение Теслы могло бы служить боевым орудием, если дополнить его динамитом. Тесла посоветовал репортеру взглянуть на вопрос более широко и осознать, что перед ним первый представитель расы автоматов (слово «робот» еще не появилось) – механических людей, которые будут выполнять утомительную работу за человека.

В своем эссе 1900 г. «Проблема увеличения энергии человечества» Тесла писал о своих плавучих автоматах: «Знания, опыт, способность принимать решения – иными словами, разум – находящегося на расстоянии оператора воплотились в машине, наделив ее способностью двигаться и разумно выполнять все свои операции… Можно сказать, что автоматы „заимствуют разум“, поскольку каждый из них, по сути, представляет собой некое продолжение находящегося на расстоянии оператора».

Тесла пошел еще дальше и предположил, что «можно создать автомат, который будет обладать „собственным разумом“ и сможет без всякого оператора полностью самостоятельно выполнять самые разные действия и операции в ответ на внешние раздражители, воздействующие на его органы чувств, как если бы он обладал интеллектом».

Теслу часто называют великим пророком нынешнего века технологий, когда реалистичные андроиды (человекоподобные роботы), такие как Надин и София, сбивают людей с толку своим умением вести беседу. На самом же деле Тесла считал, что люди – это те же автоматы, которые реагируют на внешние раздражители и именно поэтому думают и действуют. Одной из великих идей Теслы было создание «автомата, который был бы моей механической копией и реагировал бы на внешние воздействия так же, как я, но, конечно, гораздо более примитивным образом. Такой автомат, очевидно, должен обладать источником энергии, органами передвижения, управляющими органами и одним или несколькими органами чувств, способными воспринимать внешние стимулы… Не столь важно, будет ли этот автомат состоять из плоти и костей или же из дерева и стали: главное, чтобы он выполнял все функции, которые от него требуются как от разумного существа».

СМ. ТАКЖЕ Смертоносные боевые роботы (1942), «Человеческое использование человеческих существ» (1950), Беспилотные автомобили (1984)



Полученный в 1898 г. патент Николы Теслы на роботизированную лодку с беспроводным управлением, оснащенную аккумулятором, моторным винтом, рулем и фонарями. Тесла верил, что в будущем появятся разумные «телеавтоматы», которые произведут революцию в обществе.

Тик-ток. 1907

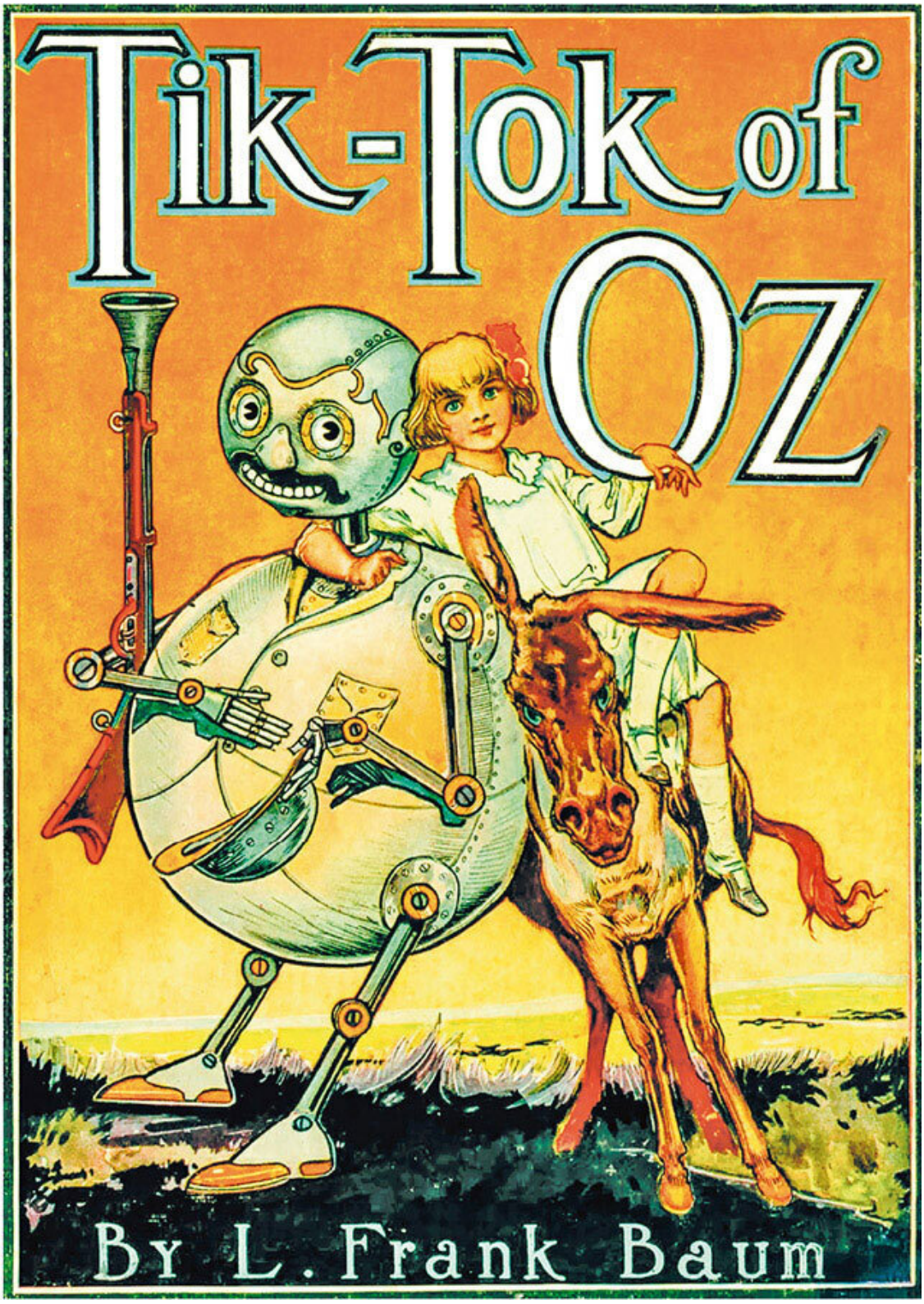
Приступая к изучению ИИ, «мы сразу же сталкиваемся с фундаментальными вопросами о природе жизни, о смерти, межполовых отношениях, работе и механизмах разума, – пишут исследователи Пол Абрам и Стюарт Кентер. – Это трудное дело требует глубокого погружения в литературу, философию и множество других областей науки и техники». Одной из первых думающих машин, описанных в художественной литературе и поднявших вопрос о тонкой грани между машинами и людьми, стал Тик-Ток – разумный медный робот, впервые появившийся в романе 1907 г. «Озма из страны Оз» американского писателя Фрэнка Баума (1856–1919). Чтобы привести в действие этого робота, требуется периодически подзаводить три его пружины, которые по отдельности отвечают за его мысли, действия и речь. Например, можно запустить мысли робота, но не действия или речь, тем самым создав изолированный «ИИ в коробке». Или же можно включить его речь, не запуская мысли: так он сможет выдавать примитивную голосовую реакцию, но будет лишен средств обработки естественного языка. Но даже когда робот полностью заведен, его модули обработки языка не вполне естественны, о чем свидетельствуют его монотонно звучащий голос и буквальная интерпретация многих вопросов и команд. По словам Баума, Тик-Ток «делает все, но не живет» и не испытывает никаких эмоций. Он не чувствовал боли от наказаний колотушками, поскольку они лишь «полировали» его медное тело.

Тик-Ток осознает свое место в мире. Например, когда его благодарят за доброту, он отвечает: «Я-машина. Я-не-добр-и-не-зол. Я-не-радуюсь-и-не-огорчаюсь»[8]. Хотя роман адресован юным читателям, Баум заставляет нас задуматься о будущем ИИ. Являются ли эмоции главным свойством, отличающим человека от машины? В какой степени литература и кино влияют на подходы к созданию ИИ и задают те ограничения, которые мы налагаем на думающие машины?

«Киборги, роботы и другие механические существа – ключевые образы для понимания технофильских и технофобских мечтаний того или иного столетия, – пишет профессор Алекс Гуди. – Они воплощают страх перед вторжением технологий, открывают перспективы технологической трансцендентальности и бросают вызов идее уникальности человека как самостоятельного субъекта».

СМ. ТАКЖЕ Медные рыцари из легенды о Ланселоте (ок. 1220), «Паровой человек в прериях» (1868), Три закона робототехники Азимова (1942), Обработка естественного языка (1954), Герметичный ящик для искусственного интеллекта (1993)

[8] Пер. С. Б. Белова.



Обложка книги «Тик-Ток из страны Оз» (1914) с иллюстрациями Джона Нила (1877–1943).

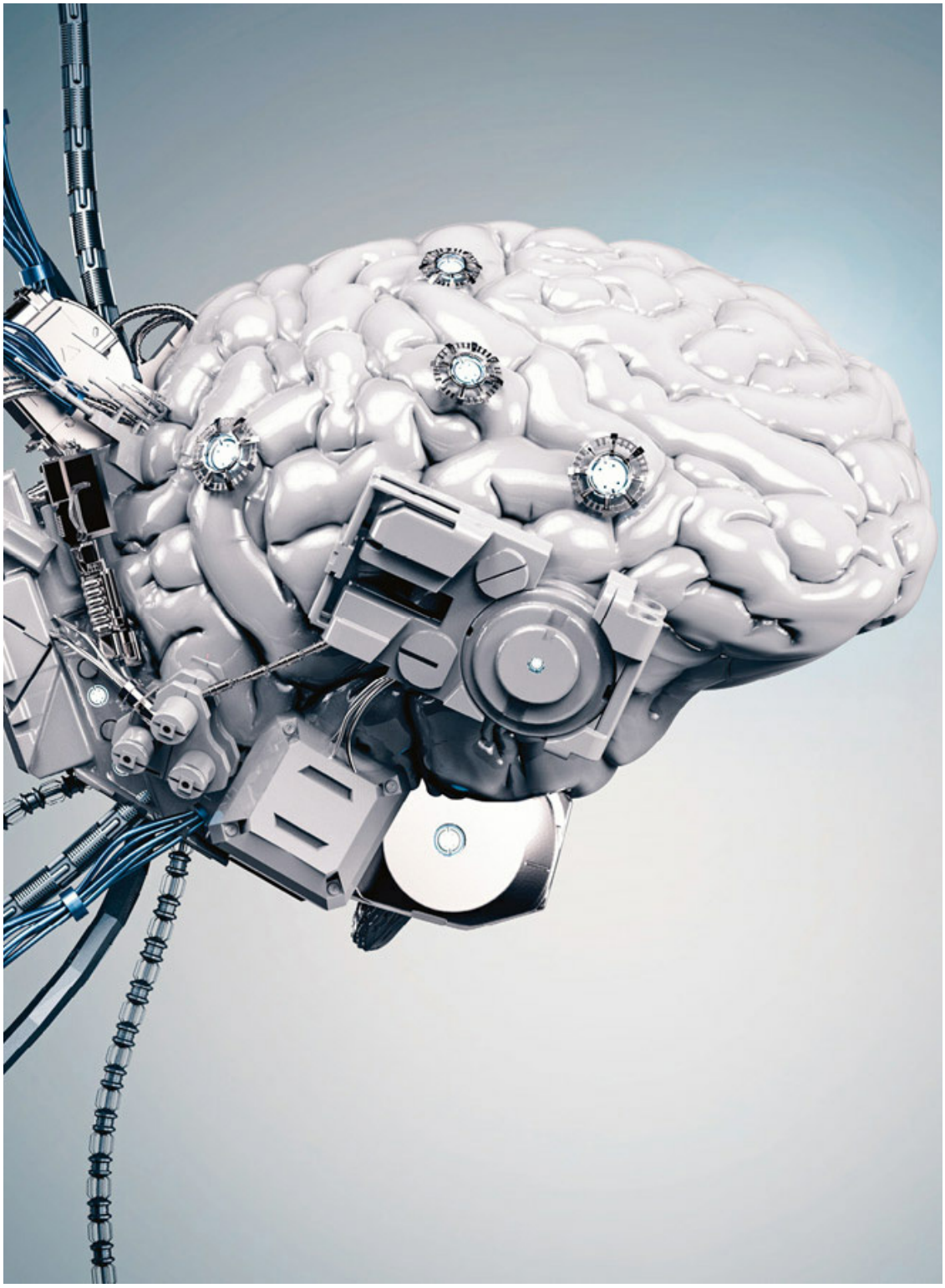
Поиски души. 1907

В своей работе 1950 г. «Вычислительная техника и интеллект» информатик Алан Тьюринг (1912–1954) писал, что, пытаясь создать машину с ИИ, мы «узурпируем» божественную «способность создавать души» не более, чем при деторождении: «Скорее мы становимся инструментами Его воли, предоставляя вместилища для душ, которые Он создает». Некоторые футурологи полагают, что, все больше узнавая о структуре мозга, человечество все ближе подходит к созданию мыслящего ИИ, который возникнет путем моделирования сознания или загрузки компонентов нашего сознания в компьютер. Эти предположения основаны на материалистической точке зрения, согласно которой сознание возникает из мозговой деятельности. С другой стороны, французский философ XVII в. Рене Декарт (1596–1650) полагал, что разум, или душа, существует отдельно от мозга. По его мнению, душа соединяется с мозгом через особый орган – шишковидное тело, которое служит воротами между мозгом и разумом.

Различные суждения о раздельном существовании души и материи отражают философию картезианского дуализма. В 1907 г., пытаясь доказать справедливость этой концепции, американский врач Дункан Макдугалл (1866–1920) помещал умирающих от туберкулеза пациентов на весы. Макдугалл рассудил, что в момент смерти, когда душа покидает тело, весы должны показывать снижение массы. По итогам экспериментов врач оценил массу души в 21 грамм. Увы, ни самому Макдугаллу, ни другим исследователям так и не удалось воспроизвести этот результат.

Более материалистический взгляд на разум и тело подкрепляется экспериментами, которые показали, что наши мысли, память и личные качества могут меняться при повреждении отдельных областей мозга. Методы мозгового картирования позволяют соотнести наши чувства и мысли с конкретными участками мозга. Вот лишь один любопытный пример: повреждение правой лобной доли мозга может вызвать внезапную страсть к изысканным ресторанам и деликатесам – так называемый *синдром гурмана*. Конечно, дуалист Декарт мог бы возразить, что повреждение мозга влияет на поведение лишь потому, что именно через мозг действует сознание. Скажем, если мы уберем руль из автомобиля, то автомобиль начнет вести себя иначе, но это не будет означать, что в нем нет водителя.

СМ. ТАКЖЕ Мельница Лейбница (1714), Трансгуманизм (1957), Жизнь в симуляции (1967), «Искусственный разум» Спилберга (2001)



Алан Тьюринг писал, что, когда люди создадут совершенные мыслящие машины, они тем самым вовсе не посягнут на божественную способность творить души – как не посягают и сейчас, когда производят на свет детей.

Распознавание текста. 1913

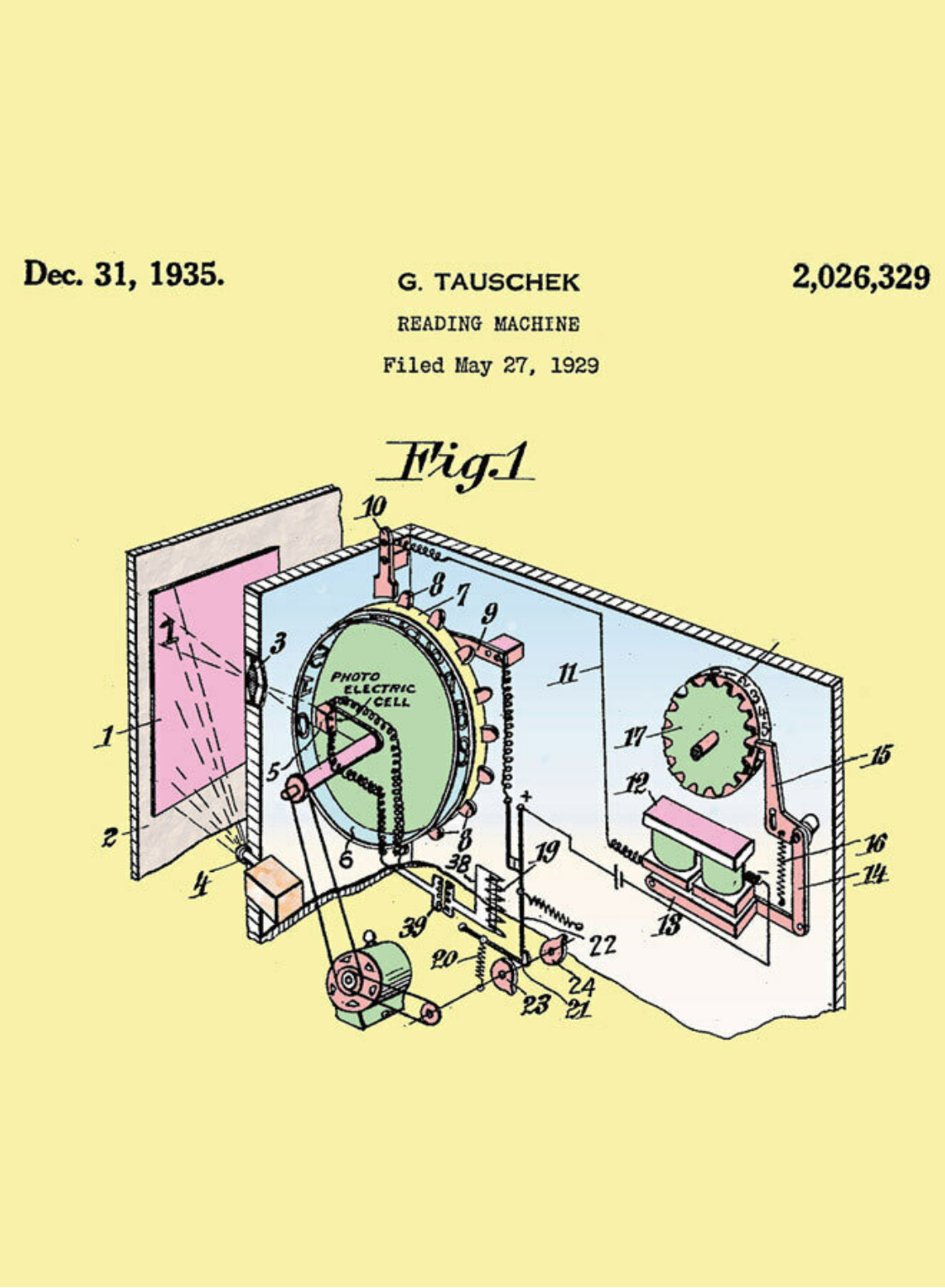
Президенту Джону Адамсу приписывают высказывание: «Единственное, что большинство людей делает лучше, чем кто-либо другой, – это разбирает собственный почерк». Попытки создания автоматизированных систем, способных читать печатный текст, имеют долгую историю. Методология *оптического распознавания символов* включает в себя целый ряд областей исследований: компьютерное зрение, искусственный интеллект, распознавание образов и т. д. Технологии позволяют преобразовывать изображения текстов (рукописных, печатных, набранных на компьютере) в машинно-кодированный текст. Машины умеют сканировать и распознавать тексты на почтовых конвертах, номерных знаках автомобилей, книжных страницах, уличных указателях и в паспортах. Эти же технологии используются для преобразования текста в речь для удобства слепых.

Один из первых изобретателей в области распознавания текста – ученый российского происхождения Эмануэль Гольдберг (1881–1970). В 1931 г. он запатентовал устройство, в котором использовались фотоэлементы, позволявшие найти нужную информацию на микрофильмированных документах. Еще раньше, примерно в 1913 г., ирландский физик Эдмунд Фурнье д'Альба (1868–1933) изобрел *оптофон* для слепых читателей: устройство сканировало текст с помощью фотодатчиков и генерировало звуки, соответствующие буквам. В 1974 г. американский изобретатель Рэй Курцвейл (р. 1948) создал читающую машину для слепых, которая сканировала тексты, напечатанные разными шрифтами, и производила речевой вывод.

Обработка распознаваемого текста включает ряд этапов: его поворот по мере необходимости, удаление шума (случайных пятен) и сглаживание краев. Система может сравнивать символы с сохраненными образцами и/или анализировать отдельные графические элементы (например, петли и линии).

Другая близкая область – *распознавание рукописного текста*. Оно может включать в себя мониторинг и анализ движений ручки для «узнавания» слова в процессе его написания. При этом обычно используется оптическое распознавание символов, но для большей точности его дополняют технологией определения наиболее вероятных слов в конкретном контексте. Для совершенствования систем распознавания печатных и рукописных текстов применяются также искусственные нейросети.

СМ. ТАКЖЕ Искусственные нейронные сети (1943), Распознавание речи (1952), Машинное обучение (1959)



На иллюстрации изображена читающая машина австрийского инженера Густава Таушека (1899–1945), на которую был выдан патент США № 2026329. На сравнивающем устройстве (диск 6) имелись прорези в виде букв и цифр. Когда изображение символа совпадало с прорезью, печатался соответствующий символ.

R.U.R. 1920

Слово «робот» вошло в английский и другие языки благодаря пьесе *R.U.R.* («Универсальные роботы Россума»), написанной в 1920 г. чешским искусствоведом и драматургом Карелом Чапеком (1890–1938).

В этой пьесе роботы изготавливаются из плоти и крови, которые, однако, замешиваются в чанах. Роботы служат человечеству как фабричные рабочие – по сути, как недорогая бытовая техника. Благодаря этому у людей появляется огромное количество свободного времени. Однако вскоре возникают дебаты о гуманности и правах подчиненных существ. Одна из главных героинь, Елена, мечтает освободить роботов. Увы, роботы, распространившись по всему миру, в конце концов уничтожают человеческую расу. Но поскольку они не обладают секретной формулой для самовоспроизводства, роботов тоже рано или поздно ждет вымирание. В конце пьесы два робота, представляющие собой своеобразных Адама и Еву для будущего планеты, влюбляются друг в друга.

Слово «робот» происходит от чешского *robota* – «каторжный труд». Пьеса стала важной вехой: она заставила людей задуматься о возможных последствиях эволюции ИИ – касающихся не только угроз для сферы труда и риска дегуманизации общества, но и безопасности человечества в целом. Где будет проведена граница между людьми и мыслящими машинами? Когда эти машины станут настолько совершенными, что потребуют для себя прав или начнут представлять угрозу? По словам Ребекки Стефофф, *R.U.R.* наводит на мысль, что «неуловимая искра человечности таится в том, как существа чувствуют и действуют, а не в том, как они были созданы».

«Философски насыщенная и противоречивая пьеса *R.U.R.* с момента своего появления была единодушно признана шедевром и стала классикой в жанре технологической антиутопии», – пишет Лучано Флориди (р. 1964), профессор философии и информационной этики Оксфордского университета. Идеи пьесы оказались настолько сильны, что к 1923 г. она была переведена более чем на тридцать языков. В 1922 г. в Нью-Йорке состоялась американская премьера постановки, после чего спектакль был сыгран там еще более сотни раз.

СМ. ТАКЖЕ «Метрополис» (1927), «Человеческое использование человеческих существ» (1950), Интеллектуальный взрыв (1965), Герметичный ящик для искусственного интеллекта (1993)

Афиша спектакля *R.U.R.*, поставленного кукольником Ремо Буфано в Нью-Йорке в 1939 г. Спектакль был профинансирован Федеральным театральным проектом (1935–1939), направленным на поддержку искусства в США в период Великой депрессии, в рамках «Нового курса» президента Франклина Рузвельта.

«Метрополис». 1927

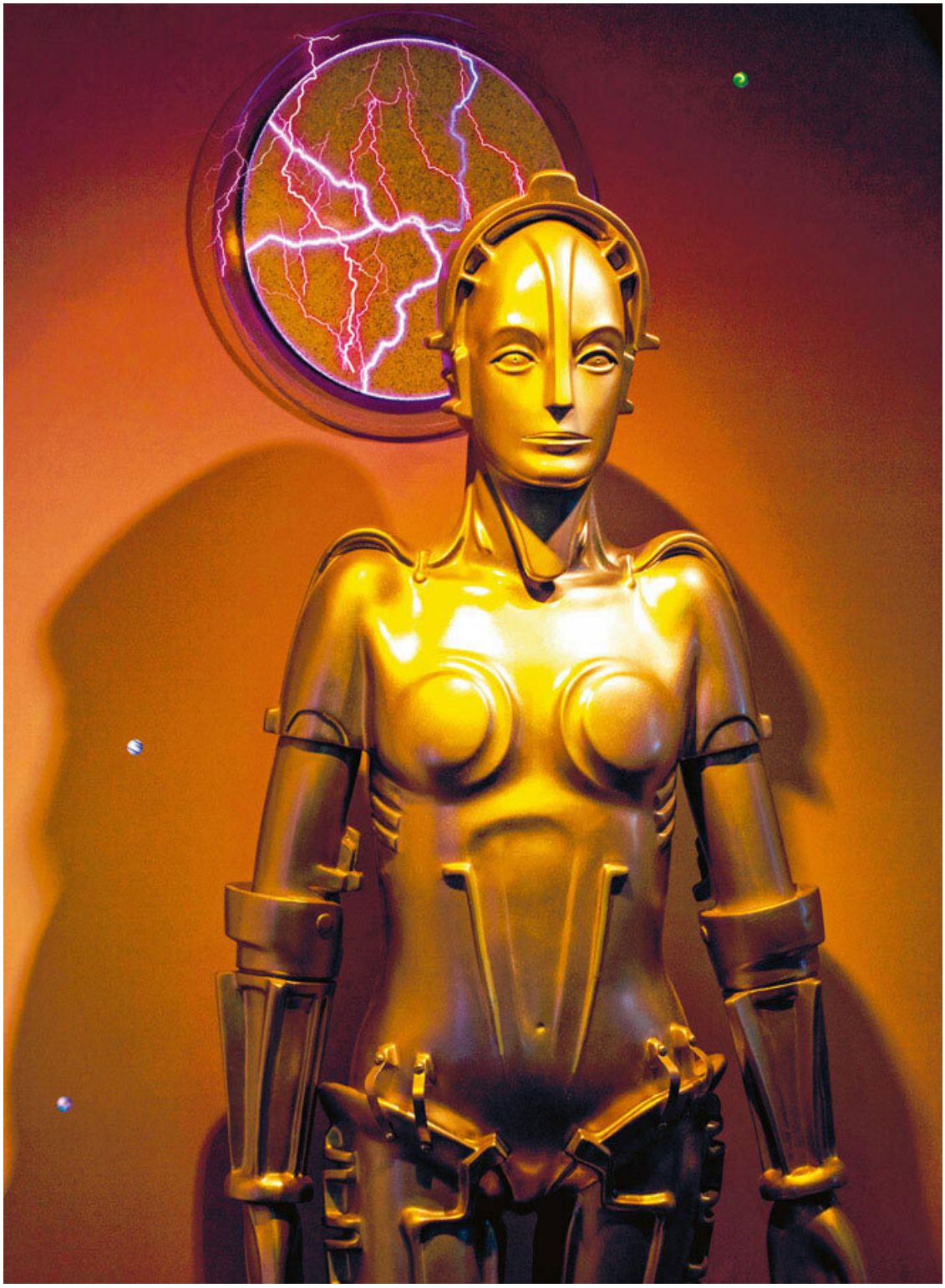
В «Метрополисе» – немом фильме 1927 г., снятом режиссером Фрицем Лангом (1890–1976) по сценарию Теи фон Харбоу (1888–1954), – один из персонажей, изобретатель Ротванг, утверждает, что его роботы никогда не устают и не ошибаются и что эти рабочие будущего вскоре не будут отличаться от людей. Действие происходит в футуристическом городе, чьи жители делятся на праздный класс, правящий городом, и низший класс, который обслуживает гигантскую подземную машину.

Героиня фильма Мария – молодая женщина, которой небезразлична судьба рабочих и трудности их жизни. По мере развития сюжета Ротванг создает робота, похожего на Марию, который по замыслу должен разрушить ее репутацию среди рабочих и отбить у них желание бунтовать. Лже-Мария, напротив, побуждает рабочих к восстанию, но позднее ее ловят и сжигают на костре. Пока она горит, человеческая оболочка тает, обнажая металлические конструкции.

Девиз этого великолепно снятого фильма – «Посредником между головой и руками должно быть сердце», а потому в нем поднимается вопрос о ключевых различиях между человеком и искусственным интеллектом. Анализируя сюжетные хитросплетения «Метрополиса», футуролог Томас Ломбардо отмечает: «По сути, робот появился в научной фантастике как условный синтез человека и машины: насколько человек становится подобен машине, будучи поглощен созданием технологии, настолько же машина уподобляется человеку, воплощая наши худшие качества и проявления… Роботы олицетворяют наш страх перед наукой и техникой, а также перед тем, кем мы можем стать».

Идеи «Метрополиса» эхом отдаются в современных научно-фантастических фильмах, таких как культовая картина об искусственных людях «Бегущий по лезвию» (1982). Конечно, чрезмерная зависимость от технологий и будущее труда в эпоху ИИ сегодня стали особенно горячими темами для дискуссий. Но «Метрополис» не утратит актуальности и в будущем, поскольку мы все чаще ошибочно принимаем ИИ-объекты за живых людей. Когда машины научатся выдавать себя за тех, кому мы доверяем и кого уважаем, или станут «смоделированными существами», в которых мы будем влюбляться, последствия могут зайти очень далеко.

СМ. ТАКЖЕ *R.U.R.* (1920), Тест Тьюринга (1950), Этика искусственного интеллекта (1976), «Бегущий по лезвию» (1982)



Робот Мария из фильма «Метрополис» в Зале славы роботов Научного центра Карнеги в Питтсбурге, штат Пенсильвания.

Подвижный робот Электро. 1939

Электро заслуживает отдельной главы в этой книге, поскольку он был провозглашен одним из первых в мире «роботов-знаменитостей» и «старейшим сохранившимся роботом Америки».

Электро был сконструирован корпорацией *Westinghouse Electric* и экспонировался на Всемирной выставке 1939–1940 гг. в Нью-Йорке, где сразу же привлек всеобщее внимание. Этот гуманоид ростом 2,1 м двигался в ответ на голосовые команды, произносил сотни слов и даже курил сигареты. Его фотоэлектрические глаза различали красный и зеленый свет. В 1940 г. у него появился компаньон – робот-собака Спарко, который мог лаять и двигаться. Посетители Всемирной выставки выстраивались в огромные очереди, чтобы посмотреть на их двадцатиминутные выступления.

Понимая, что многие могут ошибочно решить, будто робота приводит в движение сидящий внутри человек, создатели Электро намеренно вырезали в его корпусе отверстие. На самом деле механизм состоял из распределительных валов, шестерен и моторов, которые приводили в движение голову, рот и руки. Изобретатель робота, инженер Джозеф Барнетт, также использовал 78-оборотные проигрыватели, соединенные с релейными переключателями, чтобы сгенерировать словарный запас из 700 слов. Электро мог произносить фразы вроде «Мой мозг больше твоего» и реагировать на чужую речь с учетом количества услышанных слов или слогов. Например, любые три произнесенных слова активировали реле, которое останавливало движение робота. Увы, Электро и Спарко не могли уйти далеко, поскольку их контролировали операторы через специальные электрические кабели, подсоединенные к ногам.

В последующие годы Электро вдохновил множество детей на выбор инженерной профессии. Он также появился в комедийном фильме 1960 г. «Сексапильные кошечки идут в колледж». Вскоре после этого он был разобран, а его голову подарили уходящему на пенсию сотруднику *Westinghouse*. Однако в 2004 г. были обнаружены различные части робота, и Электро собрали заново.

СМ. ТАКЖЕ Робот-рыцарь Леонардо да Винчи (ок. 1495), «Паровой человек в прериях» (1868), Робот Шейки (1966), *ASIMO* и друзья (2000)

Подвижный робот Электро, представленный корпорацией *Westinghouse Electric* на Всемирной выставке 1939–1940 гг. в Нью-Йорке.

Синтез речи. 1939

Вам, вероятно, доводилось слышать синтезированный голос астрофизика Стивена Хокинга (1942–2018), который в течение многих лет использовал для общения речевой синтезатор, поскольку мышечная атрофия лишила его возможности говорить естественным образом. На самом деле технологии преобразования текста в речь служат многим полезным целям, включая чтение вслух для слабовидящих, маленьких детей и тех, кому по разным причинам трудно читать самим. Синтезированная речь также помогает разнообразным голосовым помощникам создавать впечатление человеческого интеллекта и живого общения. Новые технологии с использованием нейросетей уже позволяют имитировать естественную речь конкретных людей – а значит, в дальнейшем будет все труднее определить, действительно ли голос того, кому мы доверяем (например, делового партнера, родителя, ребенка), принадлежит самому этому человеку. Что произойдет, если кто-то «похитит» чужой голос и заставит его произносить что-либо по своему желанию?

В системах синтеза речи задействуются разные технологии. Например, система может хранить оцифрованные речевые единицы и соединять их (то есть последовательно выстраивать) во время воспроизведения. В других случаях используются акустические характеристики звуков речи (то есть форманты) – этот метод называется *формантным синтезом*. А с помощью метода *артикуляторного синтеза* можно смоделировать речевой тракт, подобный человеческому. Разумеется, простые устройства вроде говорящих часов, автомобилей, игрушек и калькуляторов хранят для воспроизведения лишь несколько предварительно записанных слов.

Существует немало сложностей, связанных с преобразованием текста в естественную и понятную речь. Например, произношение английских слов *tear, bass, read, project, desert* и многих других зависит от контекста.

Первые заметные успехи в синтезе речи связаны с именем инженера Гомера Дадли (1896–1980): он изобрел *вокодер* (кодировщик голоса), который с помощью различных электронных фильтров мог производить речь, и голосовой синтезатор *VODER* (демонстратор голосовых операций) с консолью, с помощью которой оператор создавал речь. *VODER*, имитирующий работу человеческого речевого тракта, экспонировался на Всемирной выставке в Нью-Йорке в 1939 г.

СМ. ТАКЖЕ Распознавание речи (1952), Обработка естественного языка (1954), Этика искусственного интеллекта (1976)

Голосовой синтезатор VODER, имитирующий работу человеческого речевого тракта, восхищал посетителей Всемирной выставки 1939 г. в Нью-Йорке. Для работы с устройством использовалась консоль, с помощью которой оператор создавал речь.

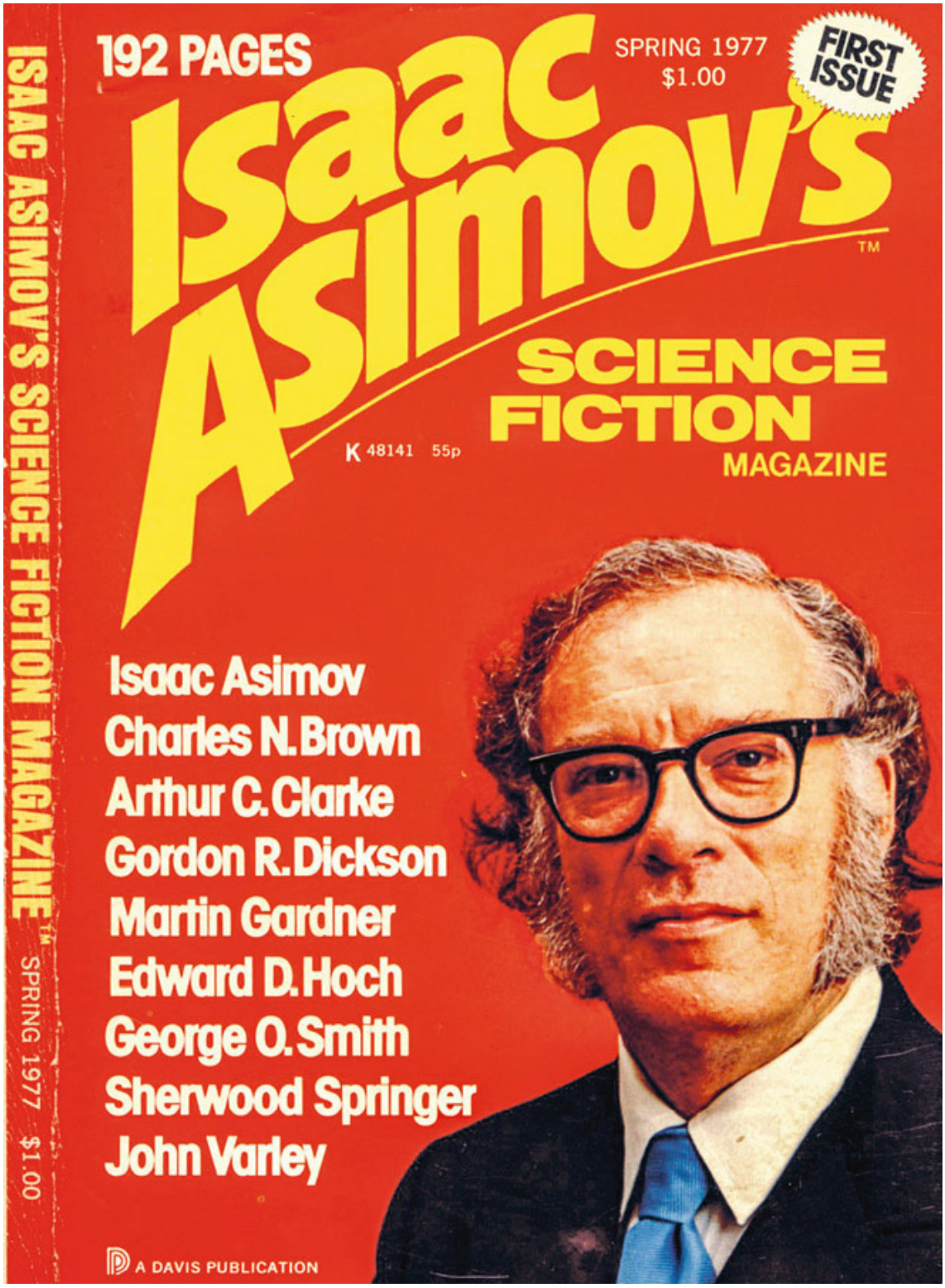
Три закона робототехники Азимова. 1942

В ближайшие десятилетия технологии ИИ и робототехника продолжат активно развиваться. Возникает вопрос: какие законы нам предстоит разработать, чтобы наши новые творения не причиняли вред людям? В 1942 г. писатель и популяризатор науки Айзек Азимов (1920–1992) в рассказе «Хоровод», повествующем о взаимоотношениях умного робота с людьми, сформулировал знаменитые три закона робототехники: 1) робот не может причинить вред человеку или своим бездействием допустить, чтобы человеку был причинен вред; 2) робот должен повиноваться всем приказам, которые дает человек, кроме тех случаев, когда эти приказы противоречат первому закону; 3) робот должен заботиться о своей безопасности в той мере, в какой это не противоречит первому или второму закону.

Позднее писатель предложил дополнительный «нулевой» закон: «Робот не может причинить вред человечеству или своим бездействием допустить, чтобы человечеству был причинен вред». Законы Азимова оказали большое влияние не только на писателей-фантастов, но и на экспертов в области ИИ. Знаменитый ученый Марвин Мински (1927–2016) отметил, что после знакомства с законами робототехники он «не переставал думать о том, как может работать сознание. Когда-нибудь мы создадим думающих роботов. Но как и о чем они будут думать? Для достижения одних целей достаточно логики, однако для других ее мало. Как создать роботов, обладающих здравым смыслом, интуицией, сознанием и эмоциями? И как со всем этим справляется наш мозг?»

Три закона робототехники полезны тем, что поднимают бесконечное количество вопросов. Какими еще законами можно дополнить триаду Азимова? Имеют ли роботы право притворяться людьми? Должны ли роботы «знать», что они роботы, и уметь объяснить, почему поступили так, а не иначе? Что, если некий террорист использует нескольких роботов, чтобы причинить вред людям, но ни один робот не будет в курсе всего плана и потому не нарушит первый закон? Как эти законы повлияют на действия армейских медицинских роботов, которым в силу невозможности помочь всем сразу придется сортировать раненых? Или на поведение беспилотного автомобиля, который должен будет решить, врезаться ли ему в играющих на дороге детей или свернуть в сторону и убить пассажира? Наконец, способен ли робот определить, что значит «причинить вред человечеству», если последствия его действий проявятся через много лет?

СМ. ТАКЖЕ Смертоносные боевые роботы (1942), Этика искусственного интеллекта (1976), «Бегущий по лезвию» (1982), Беспилотные автомобили (1984)



Писатель Айзек Азимов, сформулировавший три закона робототехники, на обложке своего научно-фантастического журнала (1977). В этом журнале был опубликован рассказ Азимова «Думай!», где автор развивает идею искусственного интеллекта.

Смертоносные боевые роботы. 1942

В XX в. роботов нередко применяли в военных операциях. Один из примеров – немецкая самоходная гусеничная мина «Голиаф», которую вермахт начиная с 1942 г. использовал на всех фронтах Второй мировой. «Голиафы» управлялись дистанционно с помощью соединительного кабеля и несли на себе взрывчатку; сами машины взрывались вместе с намеченными целями.

В наше время в качестве эффективных систем вооружения используются дроны (беспилотные летательные аппараты), оснащенные ракетами. Однако для их применения обычно требуется дистанционное участие человека: машины запрашивают «разрешение», прежде чем уничтожить цель. Первые в истории смертельные авиаудары беспилотника (*MQ-1 Predator*) были совершены в 2001 г. по целям в Афганистане. Сегодня ведутся активные дискуссии о применении в будущем *автономного* смертоносного оружия, которое сможет выявлять и атаковать военные цели без участия человека. Уже существуют автоматические оборонительные системы, в том числе машины, которые самостоятельно идентифицируют и уничтожают приближающиеся ракеты.

У военных роботов есть множество преимуществ: они не устают и не испытывают страха; они способны быстро выполнять маневры, опасные для живых пилотов; теоретически они могут сохранять жизни солдат и сокращать ущерб и число жертв среди мирного населения. Можно включить в программу робота ряд правил, например запрет на стрельбу, если точно не установлено, является ли цель вражеским бойцом, а не мирным жителем. Для большей точности можно использовать технологии распознавания лиц. Военные роботы могли бы работать плечом к плечу с солдатами, расширяя их возможности, – подобно тому как программное обеспечение и робототехника уже используются при проведении медицинских операций. Но насколько самостоятельными должны быть боевые машины? Кто будет виноват, если робот случайно атакует школу?

В 2015 г. большая группа экспертов по ИИ выступила с открытым письмом, адресованным Международной конференции по ИИ и предупреждающим об опасности военного применения не контролируемого человеком автономного наступательного оружия. По их мнению, это может привести к глобальной гонке вооружений с использованием ИИ. В числе прочих письмо подписали Стивен Хокинг, Илон Маск, Стив Возняк и Ноам Хомский.

СМ. ТАКЖЕ «Заимствованный разум» Теслы (1898), Три закона робототехники Азимова (1942), «Колосс: проект Форбина» (1970), Этика искусственного интеллекта (1976), Автономные роботы-хирурги (2016), Состязательные примеры (2018)

На иллюстрации изображен смертоносный дрон, атакующий вражеские танки после визуального распознавания и подтверждения цели с помощью ИИ.

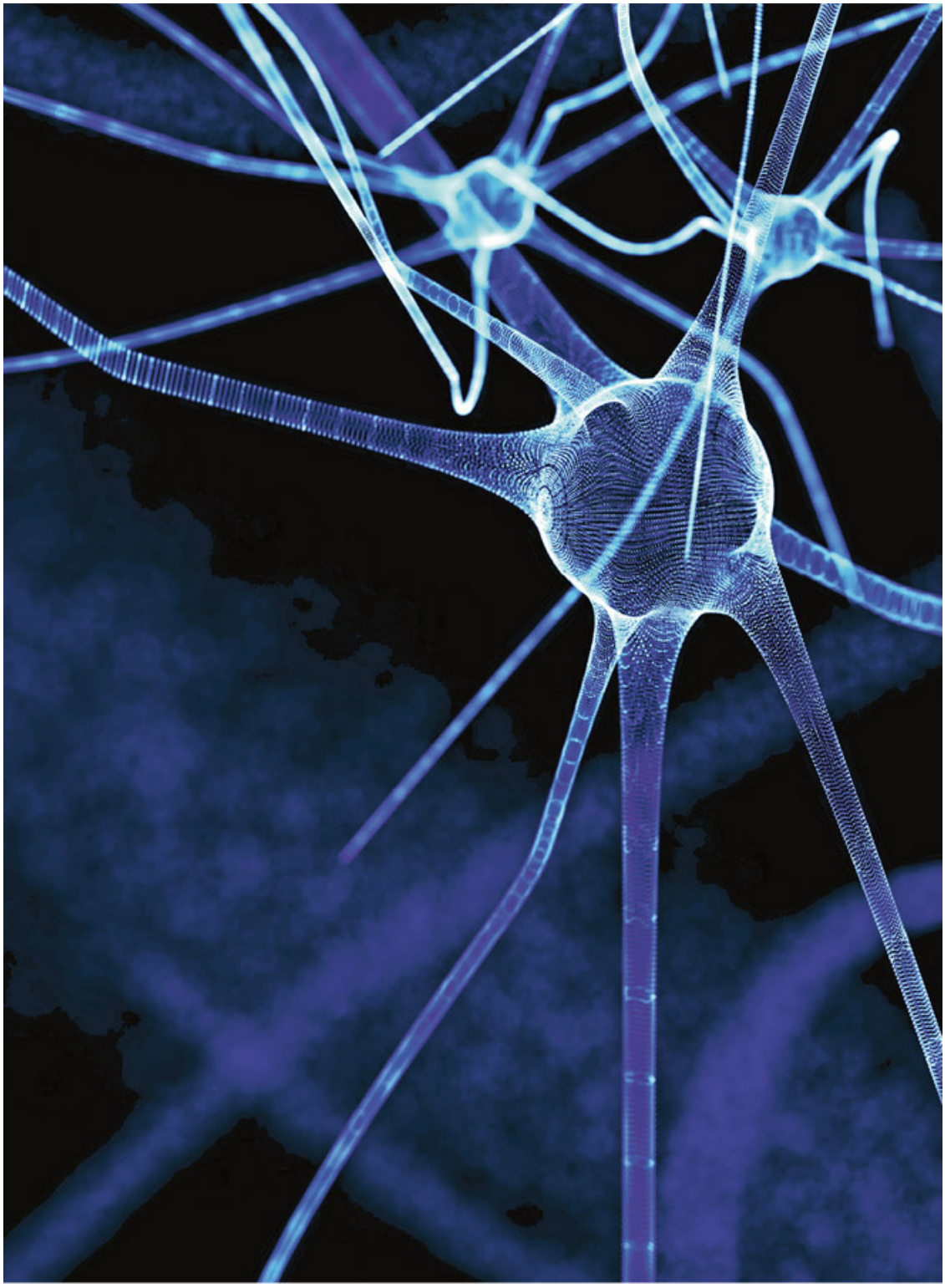
Искусственные нейронные сети. 1943

Искусственные нейросети иногда схематично изображают в виде слоеного пирога. Слои состоят из нейронов (простых вычислительных единиц), которые возбуждаются и передают возбуждение другим связанным нейронам. При этом разные весовые коэффициенты определяют, какой именно объем возбуждения должен передаваться дальше. Изначально веса и их граничные значения назначаются случайным образом, но постепенно корректируются, пока система учится выполнять задачи – например, распознавать на картинках слонов, анализируя многочисленные изображения, помеченные как «слон» и «не слон». Сегодня нейросети находят практическое применение в играх, управлении транспортом, разработке лекарств, диагностике рака по медицинским снимкам, переводах и во многих других областях.

Некоторые из базовых вычислительных моделей, применяемых в нейросетях, были рассмотрены нейрофизиологом Уорреном Маккалоком (1898–1969) и логиком Уолтером Питтсом (1923–1969) в статье 1943 г. «Логическое исчисление идей, относящихся к нервной активности», опубликованной в «Бюллетене математической биофизики». В 1957 г. Фрэнк Розенблатт (1928–1971) создал перцептронный алгоритм распознавания образов, который впоследствии был реализован на компьютере. В XXI в. ценность нейросетей возросла благодаря использованию распределенных вычислений (вычислений на разных компьютерах, объединенных в сеть) и графических процессоров.

Искусственные нейросети, созданные по принципу биологических сетей нейронов, используются в машинном обучении: компьютеры «учатся», не будучи специально запрограммированными на выполнение какой-либо задачи. Одна из проблем, связанных с нейросетями, заключается в том, что входными данными можно намеренно манипулировать, обманывая нейросеть и заставляя ее выдавать заведомо неверные ответы. И все же, размышляя о недавних примерах полезного применения нейросетей, сотрудник компании *Google*, эксперт по ИИ Джефф Дин (р. 1968) замечает: «Тот этап эволюции, на котором у животных развились глаза, стал большим шагом вперед. Теперь глаза есть и у компьютеров».

СМ. ТАКЖЕ Обучение с подкреплением (1951), Перцептрон (1957), Машинное обучение (1959), Глубокое обучение (1965), Компьютерное искусство и *DeepDream* (2015)



Искусственные нейросети устроены по принципу биологических сетей нейронов – вроде тех, по которым передаются сигналы в головном мозге.

ЭНИАК. 1946

В 1946 г. газеты пестрели восторженными отзывами об ЭНИАКе и рассуждениями о будущем думающих машин. «Механический мозг расширяет человеческие горизонты», – писала *Philadelphia Inquirer*. «Калькулятор посрамил человека», – выносила приговор *Cleveland Plain Dealer*; в той же статье говорилось о «новой эпохе в сфере человеческой мысли». Внимание СМИ к успехам в электронном воплощении мышления наряду с другими технологическими достижениями естественным образом подталкивало мир к изучению новых возможностей ИИ.

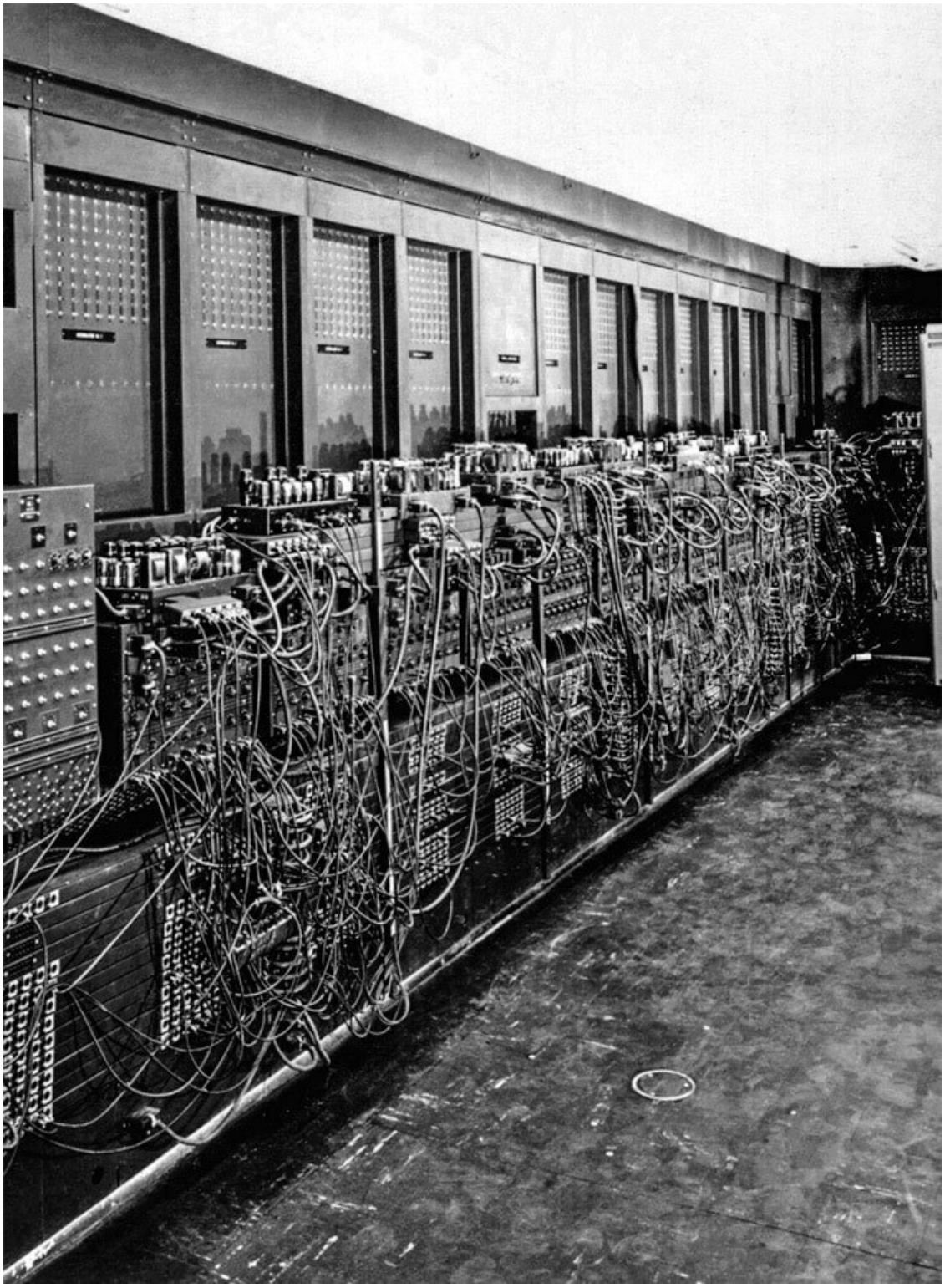
ЭНИАК («Электронный числовой интегратор и вычислитель») был построен в Пенсильванском университете американскими учеными Джоном Мокли (1907–1980) и Джоном Эккертом (1919–1995). Это устройство стало одним из первых электронных перепрограммируемых цифровых компьютеров, которые использовались для решения широкого круга вычислительных задач. Изначально ЭНИАК создавался для расчетов баллистических таблиц для армии США, однако его первое важное применение было связано с разработкой водородной бомбы.

Создание ЭНИАКа обошлось почти в 500 тысяч долларов. Аппарат был представлен в 1946 г. и почти непрерывно использовался до 2 октября 1955 г., когда был выключен окончательно. Машина содержала более 17 тысяч электронных ламп и около пяти миллионов спаянных вручную соединений. Для ввода и вывода данных использовались устройство для считывания перфокарт и карточный перфоратор. В 1995 г. группа студентов-инженеров под руководством профессора Яна ван дер Шпигеля в точности воссоздала тридцатитонный ЭНИАК на одной интегральной схеме!

Среди других важных электрических вычислительных машин 1930–1940-х гг. следует упомянуть американский компьютер Атанасова – Берри (представленный в октябре 1939 г.), немецкий Z3 (май 1941 г.) и британский «Колосс» (декабрь 1943 г.). Но, в отличие от ЭНИАКа, эти машины либо не были полностью электронными, либо имели узкое назначение.

Авторы патента на ЭНИАК (№ 3120606, зарегистрирован в 1947 г.) писали: «С приходом сложных вычислений в повседневную жизнь скорость расчетов вышла на первый план, и сегодня на рынке нет машины, способной полностью удовлетворить спрос на современные вычислительные методы… Данное изобретение призвано сократить длительность расчетов до нескольких секунд…»

СМ. ТАКЖЕ Счеты (ок. 190 до н. э.), Механический компьютер Бэббиджа (1822), «Гигантские мозги, или Машины, которые думают» (1949)



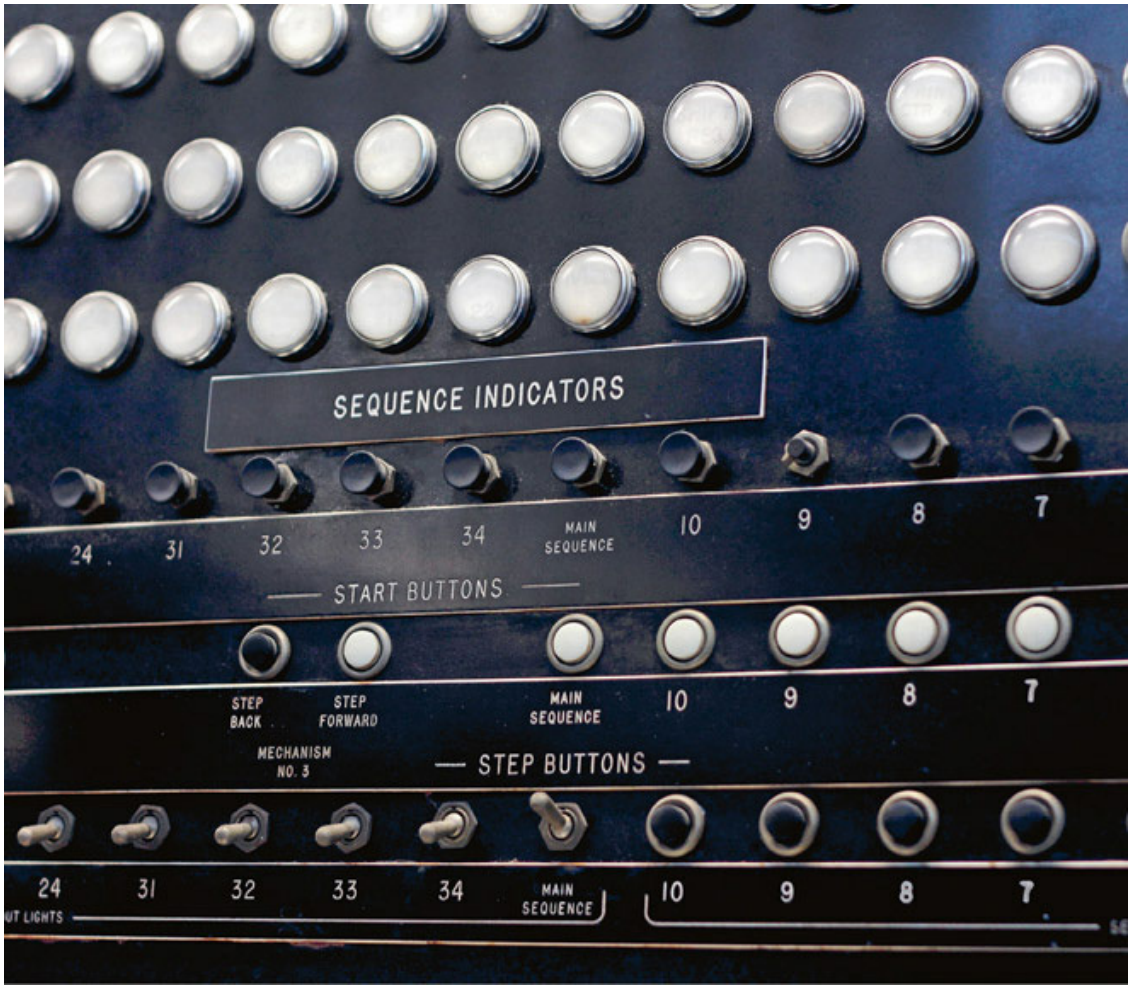
ЭНИАК был одним из первых электронных перепрограммируемых цифровых компьютеров, которые использовались для решения широкого круга вычислительных задач. Машина содержала более 17 тысяч электронных ламп.

«Гигантские мозги, или Машины, которые думают». 1949

В 1949 г. американский информатик Эдмунд Беркли (1909–1988) опубликовал книгу «Гигантские мозги, или Машины, которые думают» – вероятно, первое популярное издание о компьютерах для широкой аудитории. Книга была примечательна тем, что в ней поднимались вопросы о целесообразности употребления слов «мозг» и «думать» по отношению к компьютерам. Эти вопросы остаются открытыми и по сей день. Беркли писал: «В последнее время мы часто слышим о странных гигантских машинах, которые могут мастерски и с огромной скоростью обрабатывать информацию. Они проводят расчеты и логически рассуждают. Некоторые из них особенно умны – они способны решать гораздо больше различных задач… Им под силу проблемы, для решения которых жизнь человека слишком коротка… Примерно так выглядел бы мозг, если бы был сделан из железа и проволоки, а не из плоти и нервов. Поэтому вполне логично называть эти машины механическими мозгами».

Поразительный факт: когда Беркли писал свою книгу, электронные компьютеры были практически неизвестны широкой публике. Существовало лишь небольшое число таких «гигантских мозгов», и Беркли описывает некоторые из них. Это, например, дифференциальный анализатор № 2 в Массачусетском технологическом институте, *Mark I* в Гарвардском университете (также известный как «автоматический вычислитель, управляемый последовательностями» компании *IBM*), ЭНИАК в Школе Мура, универсальный релейный калькулятор в *Bell Laboratories* и логическая машина Калина – Беркхардта, созданная двумя студентами Гарварда. В примечаниях к изданию 1961 г. Беркли выдвигает предположение, что когда-нибудь машины смогут освоить даже интуитивное мышление: «Возможно, интуитивное мышление – это очень быстрый перебор возможных вариантов в уме вкупе с очень быстрой их оценкой: человек приходит к некоему выводу, почти не осознавая, как этот вывод был получен. В этом случае, конечно, можно запрограммировать компьютеры таким образом, чтобы они проявляли то, что мы называем интуитивным мышлением, – с той лишь разницей, что метод получения выводов нам будет известен».

СМ. ТАКЖЕ «Левиафан» Гоббса (1651), Мельница Лейбница (1714), ЭНИАК (1946), «Колосс: проект Форбина» (1970), «Назовем их искусственными пришельцами» (2015)



«Гигантский мозг». На фото – индикаторы последовательности и переключатели компьютера *Mark I* («автоматического вычислителя, управляемого последовательностями» компании *IBM*) в здании исследовательского центра Гарвардского университета.

Тест Тьюринга. 1950

Французский философ Дени Дидро (1713–1784) однажды заметил: «Если бы нашелся попугай, способный ответить на любые вопросы, я без колебаний назвал бы его разумным существом». Это подводит нас к вопросу: можно ли считать разумными созданиями, способными «думать», запрограммированные должным образом компьютеры? В 1950 г. английский информатик Алан Тьюринг попытался ответить на этот вопрос в своей знаменитой статье «Вычислительные машины и разум», опубликованной в журнале *Mind*. Он предположил, что если бы компьютер вел себя так же, как человек, то мы могли бы назвать его разумным. Далее Тьюринг предложил особый тест для оценки интеллекта компьютера. Представим, что компьютер и человек в текстовой форме отвечают на текстовые вопросы экзаменатора-человека, который не видит, кто или что ему отвечает. Если после изучения ответов экзаменатор не сможет отличить компьютер от человека, это будет означать, что компьютер успешно прошел типовую версию того, что сегодня называется тестом Тьюринга.

В наше время каждый год проводится конкурс на Премию Лёбнера, в котором программисты соревнуются в создании программ, наиболее близких к прохождению теста Тьюринга. Конечно, тест Тьюринга уже многие годы вызывает активные дискуссии, в которых поднимаются сложные вопросы. Например, если компьютер на самом деле окажется гораздо «умнее» людей, ему придется прикинуться менее умным, поскольку тест предполагает имитацию человеческого поведения. Поэтому, чтобы перехитрить жюри, разработчики часто используют обманные и неожиданные приемы: опечатки, изменение темы беседы, шутки, встречные вопросы судьям и т. д. В 2014 г. робот-собеседник, разработанный программистами из России и Украины, прошел версию теста Тьюринга, представившись тринадцатилетним украинским мальчиком Женей Густманом.

Еще одна проблема, связанная с тестом Тьюринга, заключается в том, что его результат во многом зависит от уровня подготовки судей. Однако что бы ни говорилось о способности теста выявлять «интеллект», он определенно вдохновляет на творчество программистов и разработчиков.

СМ. ТАКЖЕ «Дарвин среди машин» (1863), «Гигантские мозги, или Машины, которые думают» (1949), Обработка естественного языка (1954), Психотерапевт *ELIZA* (1964), Китайская комната (1980), Парадокс Моравека (1988)

Тест Тьюринга проверяет способность машины демонстрировать разумное поведение, неотличимое от поведения человека.

«Человеческое использование человеческих существ». 1950

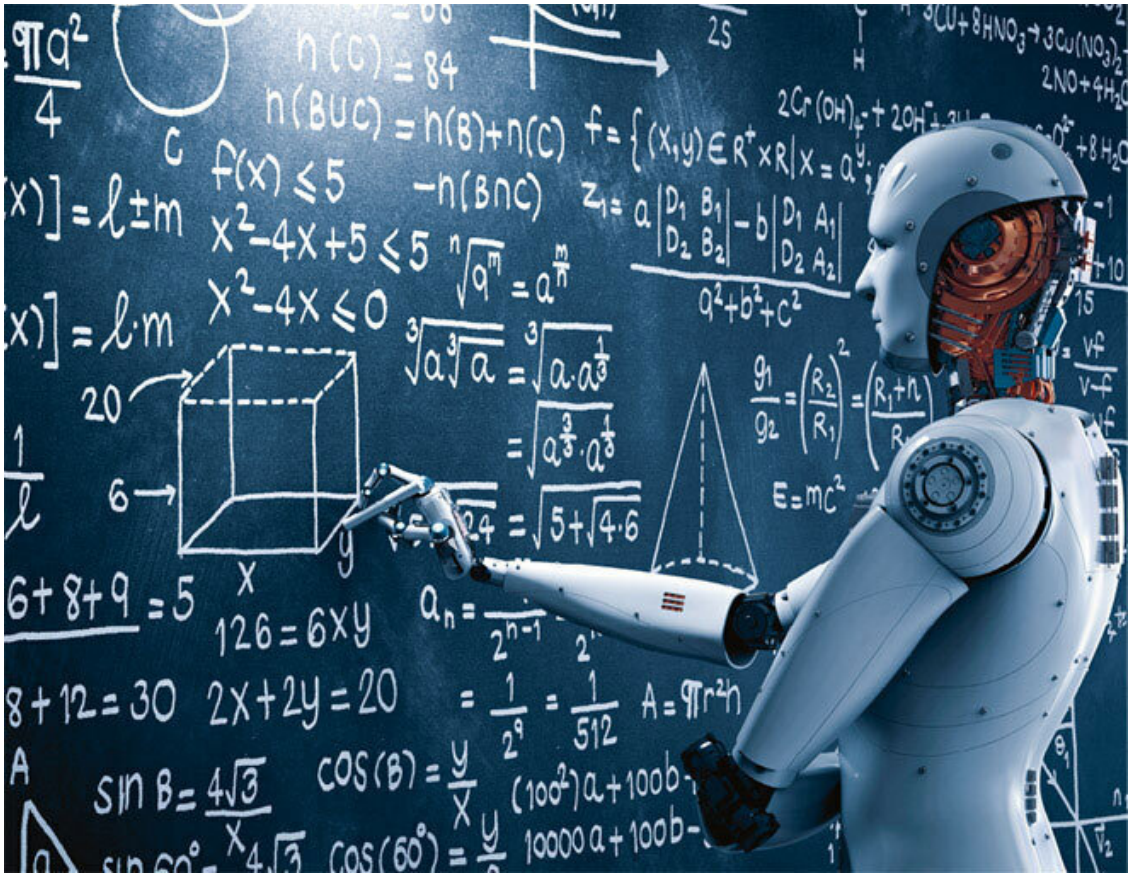
Норберт Винер (1894–1964), авторитетный американский математик и философ, был одним из отцов *кибернетики* – дисциплины, которая изучает обратную связь во многих областях человеческой деятельности и технологий. По словам эксперта по ИИ Даниэля Кревье, Винер считал, что механизмы обратной связи – это «устройства обработки информации: они получают информацию и принимают решения на ее основе. Винер высказывал гипотезу, что все разумное поведение – следствие работы механизмов обратной связи; возможно, и разум как таковой – результат получения и обработки информации».

В своей книге «Человеческое использование человеческих существ» (1950) Винер размышляет о том, каким образом люди будут сотрудничать с машинами. Его идеи, безусловно, актуальны и в нашу эпоху почти непрерывной электронной коммуникации: «Главный тезис этой книги состоит в том, что общество можно понять только путем изучения его информационных посланий и средств связи и что в будущем развитие коммуникации и средств связи между человеком и машинами, машинами и человеком, а также между машиной и машиной будет играть все более важную роль».

Винер прозорливо отмечал, что в будущем понадобится создать обучаемые машины, но вместе с тем предупреждал, как опасно делегировать процессы принятия решений машинам, мыслящим буквально: «Любая машина, построенная для принятия решений, если она не способна к обучению, будет мыслить исключительно буквально. Горе нам, если мы доверим ей руководить нами, не изучив прежде ее принципов действия и не убедившись, что она будет работать по приемлемым для нас правилам! [Машина, которая] способна учиться и принимать решения на основе обучения, вовсе не будет обязана принимать такие решения, которые приняли бы мы сами или которые будут для нас приемлемы. [Делегировать ответственность машине], способной или неспособной к обучению, – все равно что бросить ответственность на ветер и увидеть, как она возвращается вместе с бурей».

Эти предостережения актуальны и сегодня, когда многие футурологи говорят о необходимости сделать общий ИИ безопасным.

СМ. ТАКЖЕ «Дарвин среди машин» (1863), «Заимствованный разум» Теслы (1898), Дартмутский семинар по искусственному интеллекту (1956), Интеллектуальный взрыв (1965), Глубокое обучение (1965)



Норберт Винер писал, что машина, которая «способна учиться и принимать решения на основе обучения, вовсе не будет обязана принимать такие решения, которые приняли бы мы сами или которые будут для нас приемлемы».

Обучение с подкреплением. 1951

Обучение с подкреплением напоминает поведение подопытных кошек, желающих получить вознаграждение. В начале XX в. психолог Эдвард Торндайк (1874–1949) помещал кошек в клетки, из которых можно было выбраться, только наступив на рычаг. После некоторых блужданий по клетке кошка в конце концов случайно наступала на рычаг, дверца открывалась, и животное получало что-то в награду – например, еду. По мере того как кошки учились связывать нужное действие с вознаграждением, они сбегали из клетки все быстрее и в итоге нажимали на рычаг сразу после попадания в клетку.

В 1951 г. ученый-когнитивист Марвин Мински (1927–2016) и его студент Дин Эдмундс сконструировали *SNARC* – нейросетевую машину, состоящую из 3000 электронных ламп, которые имитировали 40 связанных нейронов. Мински использовал эту машину для изучения сценария, в котором условная крыса бегала по лабиринту. Когда крыса случайно совершала последовательность полезных движений и выбегала из лабиринта, связи, соответствующие этим движениям, усиливались, тем самым подкрепляя желаемое поведение и ускоряя обучение. Среди других ранних примеров устройств для обучения с подкреплением следует отметить системы для игры в шашки (1959), крестики-нолики (1960) и нарды (1992).

Как следует из этих примеров, обучение с подкреплением – это разновидность машинного обучения, которое предполагает прохождение определенных состояний в поисках вознаграждения или максимизации ожидаемого совокупного вознаграждения. «Ученик» (программный агент) совершает множество действий, чтобы выяснить, какие из них приносят наибольшее вознаграждение. Сейчас обучение с подкреплением часто совмещают с глубоким обучением, при котором задействуется крупная нейросеть, зачастую для распознавания закономерностей в данных. При обучении с подкреплением системы и машины учатся без заранее сформулированных инструкций. Это означает, что беспилотные автомобили, промышленные роботы и дроны развивают и совершенствуют свои навыки методом проб и ошибок, постепенно накапливая опыт. Однако широко применять подобный метод проблематично: он требует огромных массивов данных и тренировочных симуляций.

СМ. ТАКЖЕ Крестики-нолики (ок. 1300 до н. э.), Искусственные нейронные сети (1943), Машинное обучение (1959), Победа над чемпионом мира по коротким нардам (1979), Шашки и искусственный интеллект (1994)

Обучение с подкреплением – метод обучения программных агентов полезным действиям для максимизации общего вознаграждения. Среди ранних примеров применения метода – решение для прохождения лабиринтов, а также системы для игры в шашки, крестики-нолики и короткие нарды.

Распознавание речи. 1952

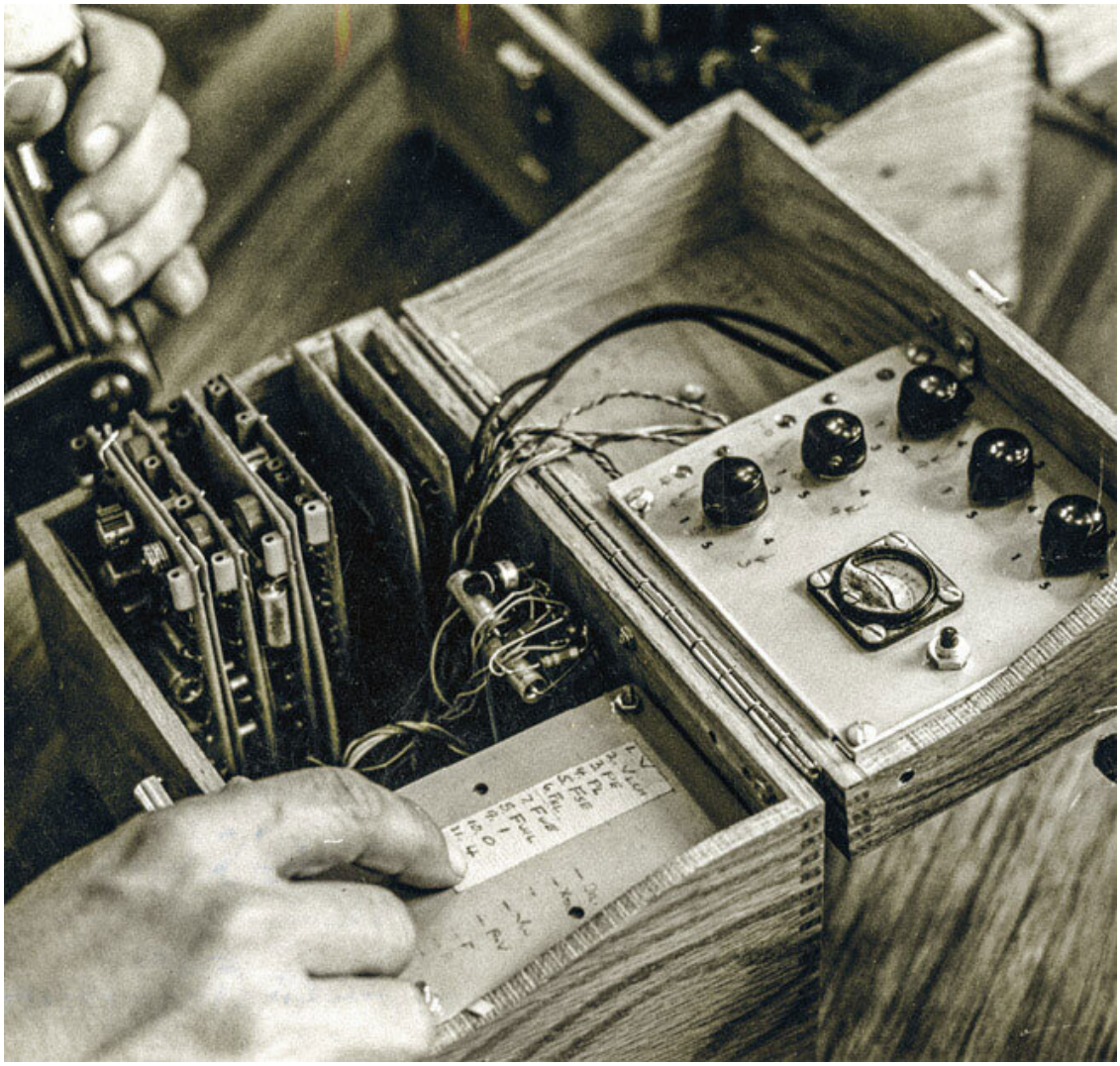
Недавно журнал *Economist* приравнял использование современных устройств с технологией распознавания речи к «произнесению магического заклинания», которое позволяет людям «управлять миром с помощью одних только слов». Это перекликается с утверждением писателя Артура Кларка: любая достаточно развитая технология неотличима от магии. «Стремительное развитие технологий обработки голоса подтверждает тезис Кларка… Скажите несколько слов в пространство – и ближайшее устройство исполнит ваше желание».

Теория и практика *машинного распознавания звучащей речи* имеют долгую историю. В 1952 г. *Bell Laboratories* разработала на основе ламповой схемы систему *AUDREY*, которая понимала произносимые вслух числа. Десять лет спустя на Всемирной выставке 1962 г. в Сиэтле была представлена машина *IBM Shoebox*: она распознавала шестнадцать слов, включая цифры от 0 до 9, и выполняла арифметические операции, если слышала такие слова, как «плюс». В 1987 г. американский производитель игрушек *Worlds of Wonder* создал куклу Джули, которая понимала несколько простых фраз и отвечала на них.

Технологии машинного распознавания речи значительно эволюционировали. Поначалу в них использовалась скрытая марковская модель – статистический метод предсказания того, соответствует ли звук слову. В наше время для достижения высокой точности распознавания применяется глубокое обучение (то есть искусственные нейросети с множеством слоев). Например, система распознавания речи может слышать звуковой поток в шумной среде и строить «догадки» о том, что говорится, определяя вероятность появления разных слов и фраз, с которыми она сталкивалась в тренировочных текстах. Специальные приложения могут располагать данными о вероятности использования той или иной фразы и определять, например, следует ли ранжировать слова «аневризма брюшной аорты» высоко или низко, с учетом того, услышаны ли они системой голосового ввода в рентгеновском кабинете или автомобильной системой, ожидающей простой команды.

Сегодня многочисленные цифровые помощники – в наших домах, автомобилях, офисах и мобильных телефонах – отвечают на голосовые команды и вопросы и пишут заметки под нашу диктовку. Речевой ввод также облегчает жизнь слабовидящим и людям с ограниченными физическими возможностями.

СМ. ТАКЖЕ Синтез речи (1939), Искусственные нейронные сети (1943), Обработка естественного языка (1954)



Устройство *IBM Shoebox* слушало, как оператор произносит цифры и арифметические команды, например: «Пять плюс три плюс восемь минус девять. Сумма».

Обработка естественного языка. 1954

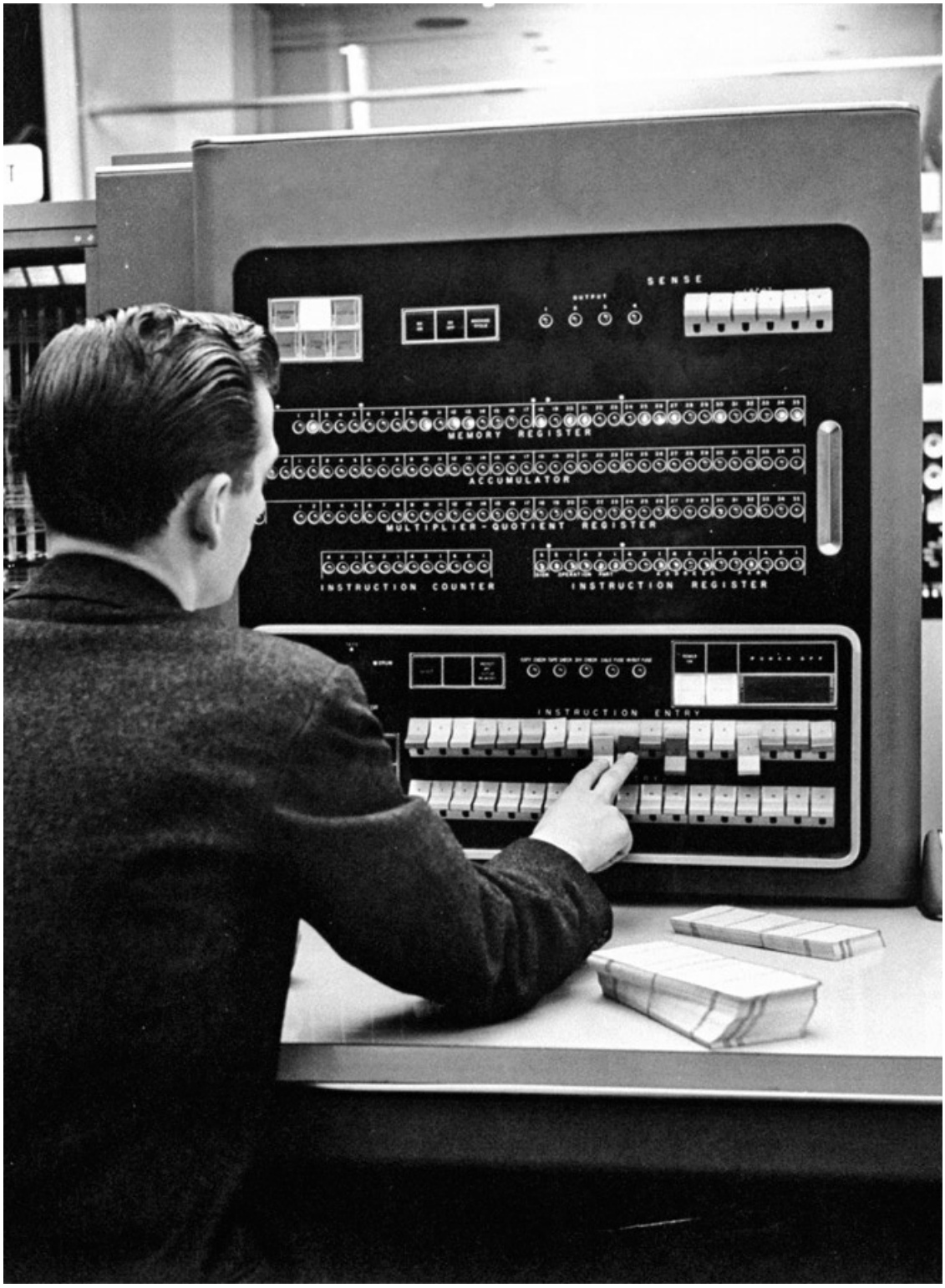
В 1954 г. вышел пресс-релиз *IBM*: «Сегодня электронный „мозг“ впервые осуществил перевод с русского языка на английский. Знаменитый компьютер 701 за несколько секунд перевел текст на удобочитаемый английский язык. Девушка, не понимавшая ни слова на языке Страны Советов, выбивала на перфокартах русские фразы». Далее сообщалось, что «„мозг“ печатает английские переводы на принтере с невероятной скоростью – две с половиной строки в секунду».

В 1971 г. информатик Терри Виноград (р. 1946) написал программу *SHRDLU*, которая переводила человеческие команды («поместить красный кубик рядом с синей пирамидой») в физические действия. В наше время под *обработкой естественного языка* понимается множество ИИ-технологий, включая распознавание речи, восприятие естественных языков (то есть понимание машиной прочитанного) и синтез речи. Одна из целей этого направления состоит в том, чтобы упростить взаимодействие между человеком и компьютером.

Поначалу для обработки естественного языка использовались сложные наборы правил, прописанных вручную; но с 1980-х гг. их заменили алгоритмы машинного обучения: машины самостоятельно осваивают правила путем анализа больших наборов данных, то есть входного языкового материала. К типичным задачам обработки языка относятся машинный перевод, ответы на вопросы («Как называется столица Франции?»), анализ тональности текста (отношение говорящего к теме) и т. д. Системы обработки естественного языка, анализирующие входные данные из текстов, аудио и видео, фильтруют спам в электронной почте, конспектируют длинные статьи и отвечают на вопросы в приложениях для смартфонов.

Обработка естественного языка сопряжена со многими трудностями. В живой речи звуки соседних слов сливаются друг с другом, и вычислительная система должна учитывать синтаксис (грамматику), семантику (значения слов) и прагматику (цель высказывания или намерение говорящего), а также контекст высказывания. Сегодня точность обработки языка удается повысить за счет широкого применения искусственных нейросетей.

СМ. ТАКЖЕ Синтез речи (1939), Искусственные нейронные сети (1943), Тест Тьюринга (1950), Распознавание речи (1952), Машинное обучение (1959), «Симбиоз человека и компьютера» Ликлайдера (1960), Психотерапевт *ELIZA* (1964), *SHRDLU* (1971), Параноик *PARRY* (1972), *Watson* участвует в телевикторине (2011)



В 1954 г., во время знаменитой демонстрации научного проекта, известного как Джорджтаунский эксперимент, русский текст был автоматически переведен на английский язык «электронным мозгом» – компьютером IBM 701, представленным на фото.

Дартмутский семинар по искусственному интеллекту. 1956

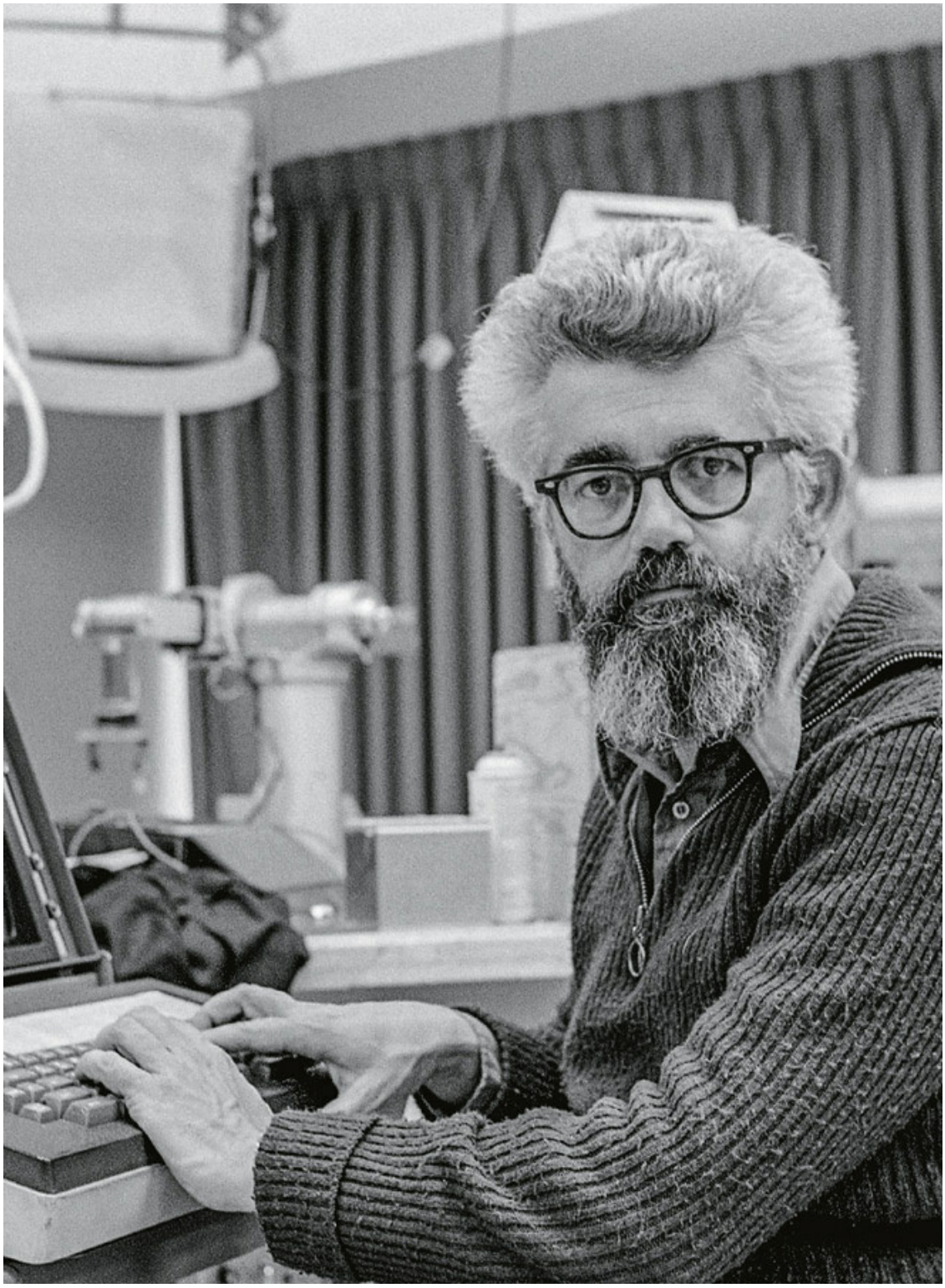
«Летом 1956 года, когда Элвис Пресли эпатировал публику вращениями бедер… а президент Дуайт Эйзенхауэр провозгласил фразу “На Бога уповаем” национальным девизом США, состоялась первая официальная конференция по ИИ», – пишет журналист Люк Дормель. Именно там, на Дартмутском летнем исследовательском семинаре по искусственному интеллекту, начал употребляться сам термин «искусственный интеллект», предложенный информатиком Джоном Маккарти (1927–2011).

Семинар был организован по инициативе Маккарти из Дартмутского колледжа, Марвина Мински (1927–2016) из Гарвардского университета, Натаниэля Рочестера (1919–2001) из *IBM* и Клода Шеннона (1916–2001) из *Bell Telephone Laboratories.* Свою заявку они сформулировали так: «Мы предлагаем исследование искусственного интеллекта сроком в 2 месяца с участием 10 человек летом 1956 г. … Исследование основано на предположении, что всякий аспект обучения или любое другое свойство интеллекта теоретически можно описать настолько точно, что машина сможет его смоделировать. Мы попытаемся выяснить, как обучить машины использовать естественные языки, формировать абстракции и концепции, решать задачи, сейчас подвластные только людям, и улучшать самих себя… Мы считаем, что вполне возможен существенный прогресс, если специально подобранная группа ученых будет работать над проблемой в течение лета». В заявке также упоминается ряд других ключевых областей исследования, в том числе «нейронные сети» и «случайность и творчество».

Во время мероприятия Аллен Ньюэлл (1927–1992) и Герберт Саймон (1916–2001) из Университета Карнеги – Меллона представили программу *Logic Theorist* для автоматического доказательства теорем с помощью математической логики. Автор книг об ИИ Памела Маккордак так пишет о Дартмутском семинаре: «Они были убеждены… что то, что мы называем мышлением, действительно может происходить вне человеческого черепа, что его можно изучить формальными и научными методами и что лучший инструмент для этого, помимо человека, – цифровой компьютер».

Отчасти из-за сложности технологий ИИ, отчасти из-за того, что участники мероприятия приезжали и уезжали в разные даты, ожидания от конференции не вполне оправдались. Тем не менее Дартмутский семинар собрал вместе самых разных исследователей, которые задавали тон в этой области в течение следующих двадцати лет.

СМ. ТАКЖЕ Искусственные нейронные сети (1943), Обработка естественного языка (1954), «Симбиоз человека и компьютера» Ликлайдера (1960)



Дартмутский летний семинар считается важной вехой в истории искусственного интеллекта. Именно там получил путевку в жизнь сам термин «искусственный интеллект», предложенный информатиком Джоном Маккарти (на фотографии он в 1974 г.).

Перцептрон. 1957

В наше время искусственные нейросети используются в бесчисленных приложениях: для распознавания образов (например, лиц), прогнозирования временных рядов (например, роста или падения цен на акции), обработки сигналов (скажем, фильтрации шума) и многого другого. Одним из важных шагов на пути к полнофункциональным нейросетям (см. главу «Искусственные нейронные сети») было создание *перцептронов*, которые разработал в 1957 г. психолог Фрэнк Розенблатт (1928–1971). В 1958 г., отчасти под влиянием его энтузиазма, газета *New York Times* провозгласила перцептрон «зародышем электронного компьютера, который, как ожидается, сможет ходить, говорить, видеть, писать, самовоспроизводиться и осознавать свое существование».

Первоначально перцептрон состоял из трех уровней связанных «нейронов» (то есть простых вычислительных единиц). Первый уровень представлял собой сетку фотоэлементов размером 20 × 20, подобную сетчатке глаза. На втором уровне находились соединительные ячейки, которые получали входные сигналы от фотоэлементов. В исходном состоянии их связи были случайными. Третий уровень содержал элемент вывода, который и называл объект, помещенный перед устройством (например, треугольник). Если перцептрон определял объект верно, исследователи усиливали электрические связи между ячейками, которые привели к распознаванию, если неправильно – ослабляли их.

Первая версия перцептрона была реализована в программном обеспечении на компьютере *IBM* 704. Для второй версии был создан перцептрон *Mark 1* – обучаемая машина, которая могла распознавать некоторые образы за счет изменения силы связей между искусственными нейронами. Математические веса кодировались в потенциометрах, а изменения весов в процессе обучения производились с помощью электродвигателей. Высказывались предположения, что устройство сможет выполнять широкий спектр задач распознавания образов, однако ажиотаж вокруг этой скромной модели превосходил ее возможности, и надежды не оправдались. В 1969 г. Марвин Мински (1927–2016) и Сеймур Паперт (1928–2016) из Массачусетского технологического института опубликовали книгу «Перцептроны», в которой описали недостатки простых перцептронов, и интерес к этой новой области машинного обучения угас. Позднее, однако, стало ясно, что конфигурации искусственных нейронов с большим количеством слоев могут иметь колоссальную ценность и применяться в самых разных сферах.

СМ. ТАКЖЕ Искусственные нейронные сети (1943), Машинное обучение (1959), Глубокое обучение (1965)

Первая версия перцептрона была реализована в программном обеспечении на компьютере *IBM* 704, изображенном на этой фотографии 1957 г. *IBM* 704 был одним из первых серийных компьютеров для выполнения арифметических действий с плавающей запятой.

Трансгуманизм. 1957

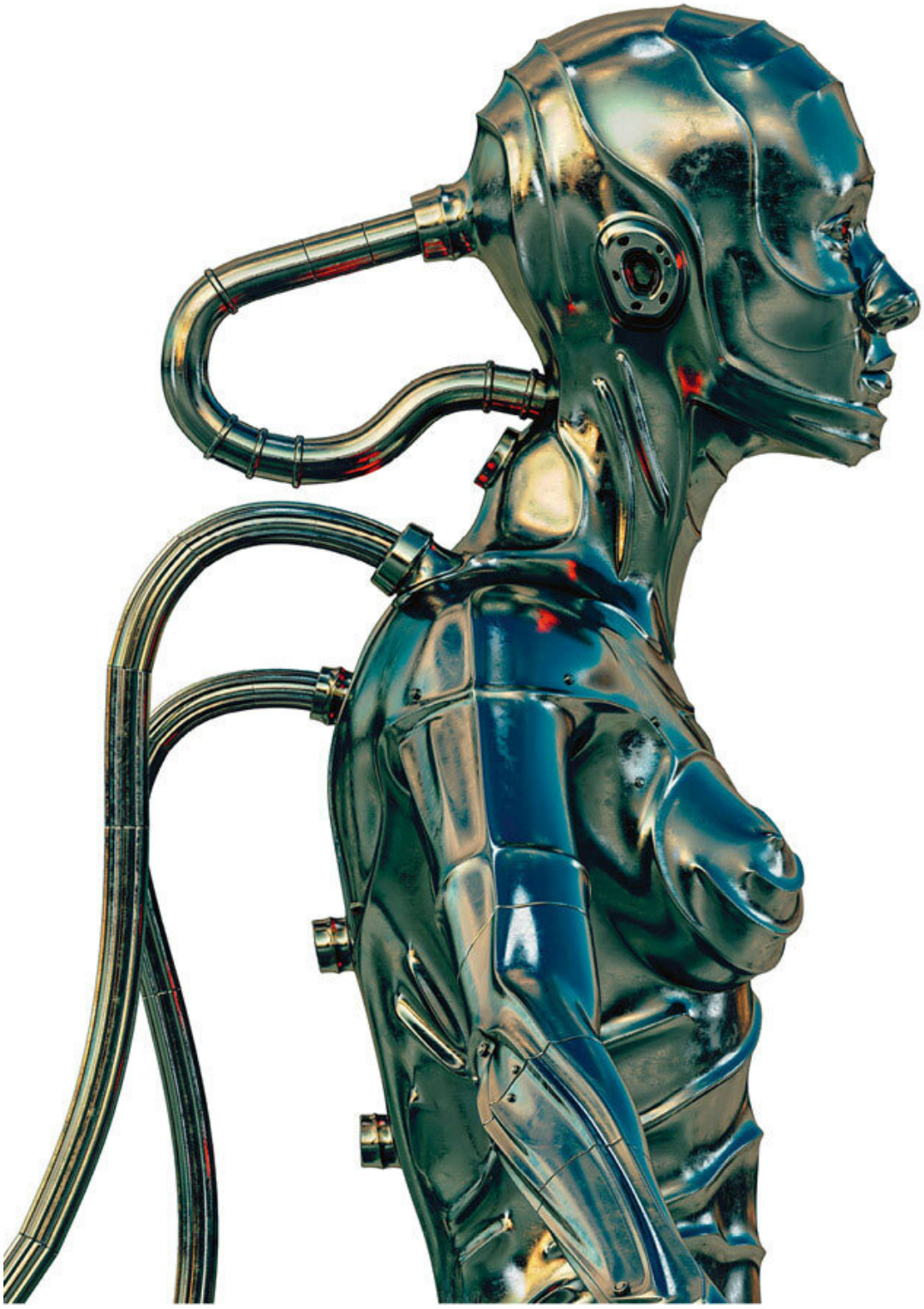
«Появление искусственного интеллекта, вероятно, станет самым значительным событием в истории человеческого вида, – пишет философ-трансгуманист Золтан Иштван (р. 1973). – Идея, конечно, не в том, чтобы позволить искусственному интеллекту развиваться бесконтрольно, вне нашего поля зрения, а в том, чтобы нам самим стать киборгами, отчасти машинами, чтобы подключаться к искусственному интеллекту напрямую, когда потребуется».

Термин «трансгуманизм» ввел в употребление биолог Джулиан Хаксли (1887–1975) в своей книге «Новые мехи для нового вина», опубликованной в 1957 г. Он высказал предположение, что «человеческий вид может… превзойти себя… реализовав новые возможности человеческой природы ради нее самой… Человеческий вид окажется на пороге нового способа существования, столь же отличного от нашего, как наш отличается от способа существования синантропа. И таким образом наш вид осознанно исполнит свое истинное предназначение».

Современная концепция трансгуманизма, разделяемая философом-футурологом Максом Мором (р. 1964) и многими другими учеными, предполагает использование технологий для расширения умственных и физических возможностей человека. Идея состоит в том, что когда-нибудь мы наверняка станем «постчеловеками» или даже обретем бессмертие благодаря генетическим манипуляциям, робототехнике, нанотехнологиям, компьютерам или переносу сознания в виртуальные миры – а также благодаря полному пониманию биологических процессов старения. Мы уже учимся использовать интерфейсы «мозг – компьютер», чтобы соединяться с продвинутым ИИ для расширения наших когнитивных способностей, и чем глубже мы понимаем биологический механизм старения, тем ближе мы к бессмертию.

Если бы ваше тело или разум могли существовать вечно, что осталось бы от вашего «я»? Мы все меняемся благодаря своему опыту – но эти изменения происходят постепенно, а значит, сейчас вы почти тот же человек, каким были год назад. Однако, если ваше обычное или улучшенное тело будет непрерывно существовать тысячу лет, за это время ваша психика значительно изменится, и в вашем теле, возможно, окажется совершенно другой человек. Этот тысячелетний человек может совсем не походить на вас. Вы словно перестанете существовать. Не будет определенного момента смерти, в который ваше существование прекратится, но все же вы будете медленно исчезать на протяжении тысячелетий, подобно замку из песка, размываемому океаном времени.

СМ. ТАКЖЕ Мельница Лейбница (1714), Поиски души (1907), «Симбиоз человека и компьютера» Ликлайдера (1960), Жизнь в симуляции (1967), «Искусственный разум» Спилберга (2001)



Американский ученый и экономист Ёсихиро Фрэнсис Фукуяма (р. 1952) назвал трансгуманизм, который предполагает использование технологий для расширения умственных и физических возможностей человека, самой опасной идеей в мире.

Машинное обучение. 1959

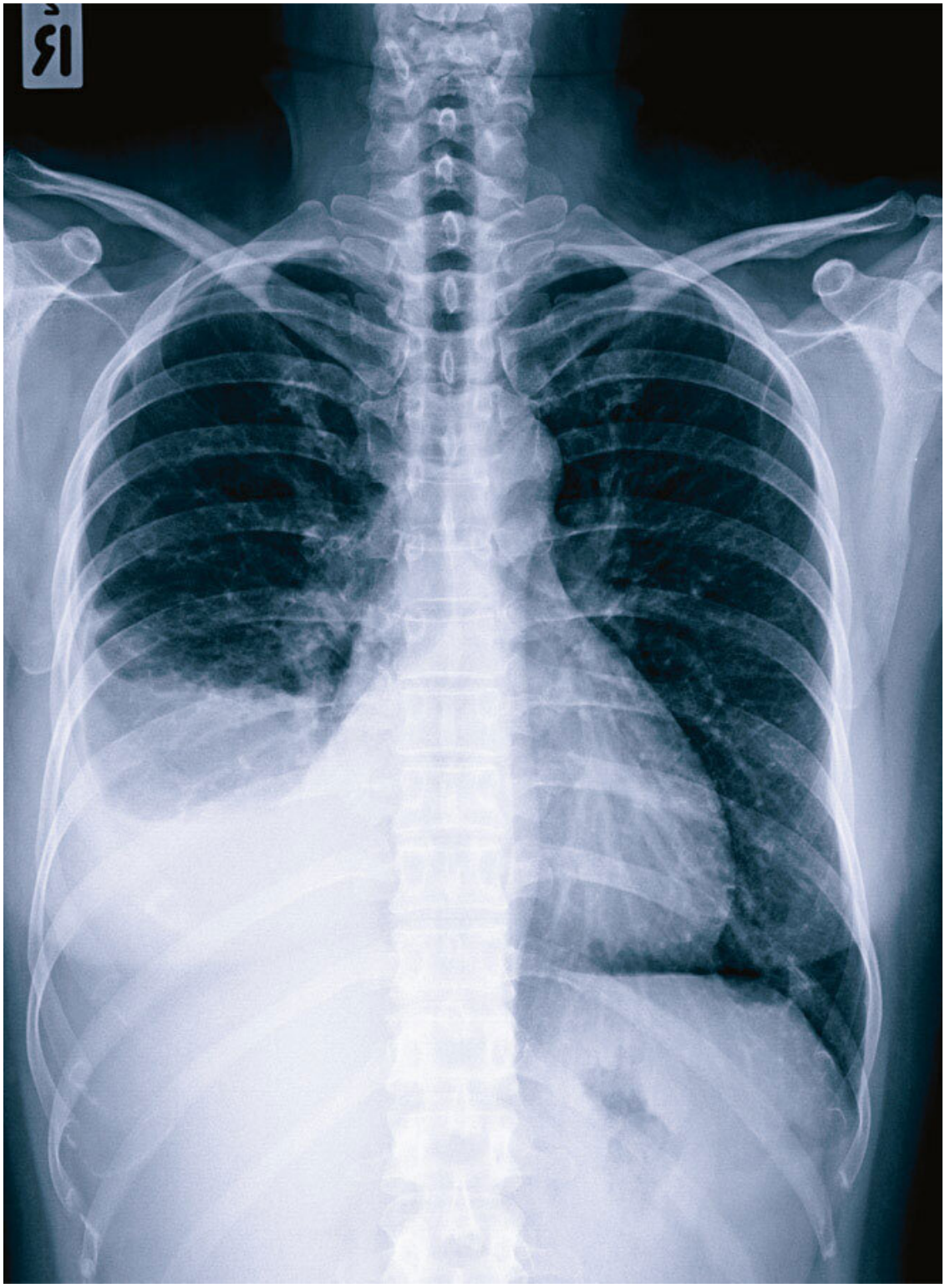
Эксперт по ИИ Артур Ли Сэмюэл (1901–1990) считается одним из первых, кто употребил термин «машинное обучение». Термин вошел в обиход благодаря статье Сэмюэла «Некоторые исследования в области машинного обучения на примере игры в шашки», опубликованной в *IBM Journal of Research and Development* в 1959 г. В статье объяснялось, что «программирование компьютеров для их обучения на опыте» может во многом устранить необходимость в инструкциях для выполнения конкретных задач.

Сегодня машинное обучение стало одним из основных методов вычислений и инструментов искусственного интеллекта. Оно играет важную роль в технологиях компьютерного зрения, распознавания речи, автономных роботов, беспилотных автомобилей, распознавания лиц, фильтрации спама, оптического распознавания символов, рекомендаций товаров, выявления риска раковых заболеваний, обнаружения утечек данных и т. д. Многие формы машинного обучения требуют больших массивов входных данных для «тренировки» системы.

При *обучении с учителем* алгоритмы получают образцы размеченных данных. На их основе система строит прогнозы для работы с новыми данными. Например, система получает набор из 100 тысяч изображений львов и тигров, которые правильно размечены людьми как «львы» или «тигры». Алгоритм обучения с учителем должен научиться отличать львов от тигров на изображениях, которых он раньше не видел. В *машинном обучении без учителя* используются неразмеченные данные, и система должна самостоятельно выявлять в них скрытые закономерности, например, определять, что тридцатилетние женщины, которые перестают покупать консервы с тунцом, могут быть беременны, а значит, становятся целевой аудиторией для рекламы детских товаров.

Методы машинного обучения не защищены от ошибок – например, если входные данные необъективны, неверны или намеренно сфальсифицированы. Поэтому нельзя слишком полагаться на автоматизированные методы при принятии решений о выдаче кредита, приеме на работу или одобрении условно-досрочного освобождения. Этот принцип актуален для самых разных областей, где в принятии решений участвуют машины.

СМ. ТАКЖЕ Искусственные нейронные сети (1943), Обучение с подкреплением (1951), Обработка естественного языка (1954), Представление знаний (1959), Глубокое обучение (1965), Генетические алгоритмы (1975), Роевой интеллект (1986), Состязательные примеры (2018)



В 2017 г. ученые из Стэнфорда разработали алгоритм машинного обучения, способный диагностировать пневмонию лучше рентгенологов. На рентгеновском снимке грудной клетки виден плевральный выпот в правом легком.

Представление знаний. 1959

«Чтобы система была интеллектуальной, – пишет информатик Нильс Нильссон, – она должна обладать знаниями о своем мире и уметь делать выводы из этих знаний или, по крайней мере, действовать исходя из них. Как у людей, так и у машин должны быть способы представления необходимых знаний во внутренних структурах, будь то кодирование в белке или кодирование в кремнии». В наше время внимание к ИИ по большей части сосредоточено на машинном обучении и статистических алгоритмах, например для распознавания образов. Однако логическая модель представления знаний по-прежнему играет большую роль во многих областях.

Представление знаний – это область исследований ИИ, связанная с хранением и обработкой информации таким образом, чтобы компьютерные системы могли эффективно ее использовать, например для постановки медицинских диагнозов и выработки юридических рекомендаций, а также для обеспечения работы интеллектуальных диалоговых систем, таких как Сири в *iPhone* или Алекса в *Amazon Echo*. В частности, представление знаний может осуществляется с помощью *семантической сети*, отражающей семантические (то есть смысловые) отношения между понятиями. Семантические сети часто принимают форму графов, вершины которых соответствуют понятиям, а ребра (соединительные линии) указывают на семантические отношения между ними. Разные методы представления знаний находят применение в автоматических рассуждениях, в том числе при автоматическом доказательстве математических теорем.

Среди первых успехов в области представления знаний можно назвать «Универсальный решатель задач» – компьютерную программу, созданную в 1959 г. Алленом Ньюэллом (1927–1992), Гербертом Саймоном (1916–2001) и их коллегами для анализа целей и решения простых задач общего характера (таких, как «Ханойская башня»). А в 1984 г. Дуглас Ленат (р. 1950) инициировал проект *Cyc* с участием множества аналитиков, которые документировали знания и факты из различных областей, чтобы помочь ИИ-системам рассуждать подобно человеку (в частности, механизм логического вывода *Cyc* использует дедукцию и индукцию). В настоящее время исследователи в области представления знаний занимаются многими вопросами, в том числе обеспечивают обновление базы знаний и ищут способы устранить неопределенность в системах представления знаний.

СМ. ТАКЖЕ «Органон» Аристотеля (ок. 350 до н. э.), Ханойская башня (1883), Перцептрон (1957), Машинное обучение (1959), Экспертные системы (1965), Нечеткая логика (1965)

Экспертная система *MYCIN* использовала ИИ для выявления бактерий, вызывающих тяжелые инфекции, и для рекомендации методов лечения. В *MYCIN* был простой механизм вывода и база знаний примерно из 600 правил. На иллюстрации – бактерии *Streptococcus pneumoniae*, вызывающие менингит.

«Симбиоз человека и компьютера» Ликлайдера. 1960

В 1960 г. увидела свет основополагающая статья психолога и информатика Джозефа Ликлайдера (1915–1990) под названием «Симбиоз человека и компьютера». Он начинает с описания симбиотических связей инжира и осы *Blastophaga*: оса опыляет инжир, а ее яйца и личинки, в свою очередь, питаются плодами дерева. Ликлайдер предположил, что люди и компьютеры точно так же могли бы создавать симбиотические связи. В первые годы симбиоза люди ставили бы цели и формулировали гипотезы, а компьютеры прокладывали бы путь к новым открытиям. Некоторые проблемы, по словам Ликлайдера, «просто невозможно сформулировать без помощи вычислительной машины».

Ликлайдер не говорил о компьютеризованных существах, которые могут вытеснить людей; его идеи ближе к концепциям Норберта Винера (1894–1964), основанным на тесном взаимодействии между людьми и машинами. В своей статье он объясняет: «Есть надежда, что… человеческий мозг и вычислительные машины будут очень тесно связаны друг с другом, и это партнерство сможет мыслить так, как никогда не мыслил ни один человеческий мозг, и обрабатывать данные так, как и близко не способны их обрабатывать современные машины».

Ликлайдер также рассуждает о «мыслительных центрах», которые, как он верил, будут выполнять функции традиционных библиотек. По его предположению, для симбиоза потребуются методы обработки естественного языка.

В своей статье Ликлайдер допускает, что «электронные или химические „машины“ превзойдут человеческий мозг в большинстве функций, которые сегодня кажутся нам исключительно его вотчиной», и приводит в качестве примеров игру в шахматы, решение задач, распознавание образов и доказательство теорем. Он уточняет, что «компьютер станет машиной для статистического анализа, будет применяться в теории принятия решений и теории игр, давать элементарные оценки предполагаемых вариантов действий… Наконец, он будет ставить диагнозы, выявлять закономерности и взаимосвязи – в той мере, в какой это себя оправдает…».

В наше время, по прошествии шестидесяти лет, вопросы о возможном союзе человеческого разума и искусственного интеллекта, поднятые в статье Ликлайдера, все еще актуальны. Когда наша связь с машинами станет еще теснее, чем сегодня, будет ли такая симбиотическая личность по-прежнему считаться «человеком»? И захочет ли эта личность вообще отсоединяться от компьютера?

СМ. ТАКЖЕ «Дарвин среди машин» (1863), Обработка естественного языка (1954), Трансгуманизм (1957)

Джозеф Ликлайдер писал: «Есть надежда, что… человеческий мозг и вычислительные машины будут очень тесно связаны друг с другом, и это партнерство сможет мыслить так, как никогда не мыслил ни один человеческий мозг».

Психотерапевт ELIZA. 1964

ELIZA – компьютерная программа, которая отвечает на текстовые сообщения, набранные на естественном языке, имитируя диалог между пользователем и психотерапевтом. Она была создана в 1964 г. информатиком Джозефом Вейценбаумом (1923–2008) и стала одним из первых и наиболее убедительных виртуальных собеседников (диалоговых речевых симуляторов). Вейценбаум был потрясен и расстроен тем, какие глубокие эмоции и какую личную информацию раскрывали некоторые люди при общении с *ELIZA*: они словно считали ее реальным человеком, способным к сопереживанию.

Программа *ELIZA* была названа в честь Элизы Дулитл из пьесы «Пигмалион», написанной в 1912 г. ирландским драматургом Джорджем Бернардом Шоу. В пьесе профессор Генри Хиггинс учит необразованную девушку Элизу правильно говорить и тем самым убедительно изображать леди из высшего общества. Аналогичным образом *ELIZA* Вейценбаума была запрограммирована отвечать на ключевые слова и фразы, создавая иллюзию реального человеческого сопереживания. Некоторые исследователи даже полагали, что эта программа действительно способна помочь людям с некоторыми психологическими проблемами.

Наблюдая за разговорами людей с *ELIZA*, Вейценбаум все больше беспокоился от того, как они попадают в зависимость от компьютеров и как легко их можно обмануть. В 1966 г. в статье о своей программе он писал: «Машины с ИИ делают удивительные вещи, которых часто достаточно, чтобы поразить даже самых искушенных наблюдателей. Но как только с программы срывается маска, как только объясняются принципы ее работы… ее магия рассыпается в прах; она оказывается простым набором процедур… Наблюдатель думает: „Я и сам мог бы это написать“. С этой мыслью он убирает программу с полки „умное“ и перекладывает на полку для безделушек… Цель этой статьи в том, чтобы вызвать именно такую переоценку программы, которая будет описана далее. Немногие из когда-либо созданных программ нуждались в этом больше».

Сегодня виртуальные собеседники (чат-боты) часто используются в диалоговых системах техподдержки и для различных видов онлайн-помощи и психотерапии. Кроме того, их встраивают в некоторые игрушки, а также применяют для консультаций клиентов онлайн-магазинов или в качестве рекламных агентов.

СМ. ТАКЖЕ Тест Тьюринга (1950), Обработка естественного языка (1954), Параноик *PARRY* (1972), Этика искусственного интеллекта (1976)

Портрет Элизы Дулитл, продавщицы цветов из пьесы Бернарда Шоу «Пигмалион», написанный художником Уильямом Брюсом Эллисом Ранкеном (1881–1941). Программа *ELIZA* была названа в честь мисс Дулитл, поскольку та научилась убедительно изображать утонченного и образованного человека, улучшив свои речевые навыки.

Распознавание лиц. 1964

Системы распознавания лиц идентифицируют людей по изображениям или видеозаписям, зачастую сравнивая черты лица (например, относительное расположение глаз и носа) с образцами из базы данных. В ряде современных систем используются трехмерные датчики для сбора информации и повышения точности распознавания при разном освещении и углах обзора, а в некоторых смартфонах для аутентификации пользователя применяется инфракрасная подсветка лица. Точному распознаванию лиц препятствует множество факторов, например аксессуары, такие как шляпы и солнцезащитные очки, или даже макияж; но уже сегодня в некоторых обстоятельствах алгоритмы распознают лица лучше, чем люди. Истоки технологии распознавания лиц можно найти в Англии XIX в., где в 1852 г. на смену негуманным методам клеймения преступников пришла система фотографирования. Это позволяло следить за заключенными и сообщать о них другим полицейским участкам в случае побега. Одним из основоположников более совершенных методов распознавания лиц был математик и информатик Вудро Бледсоу (1921–1995), который работал над этой проблемой в 1964 г. В то время он отмечал, что задачу осложняют повороты и наклоны головы, разное освещение, выражение лица, возраст и т. д. Бледсоу и другие пионеры этой области, как правило, опирались на активное взаимодействие человека с компьютером: людям приходилось вручную зарисовывать схемы лиц по фотографиям с помощью графического планшета (блокнота).

Уже много лет в системах распознавания лиц используются различные технологии, включая алгоритм *Eigenface*, скрытые марковские модели и динамическое сопоставление ссылок. Как отмечает специалист по технологиям Джесси Дэвис Уэст, распознавание лиц сегодня применяется в нескольких важных сферах: «Правоохранительные органы прибегают к распознаванию лиц для обеспечения безопасности жителей. Розничные магазины предотвращают преступность и насилие. Аэропорты повышают удобство и безопасность путешественников. А производители мобильных телефонов используют технологию распознавания лиц, чтобы предложить клиентам биометрическую защиту нового уровня». И все же остается вопрос: не указывает ли все это на тревожный поворотный момент в истории цивилизации, после которого люди уже не смогут сохранять анонимность в публичном пространстве?

СМ. ТАКЖЕ Распознавание текста (1913), Распознавание речи (1952), Робот *AIBO* (1999)

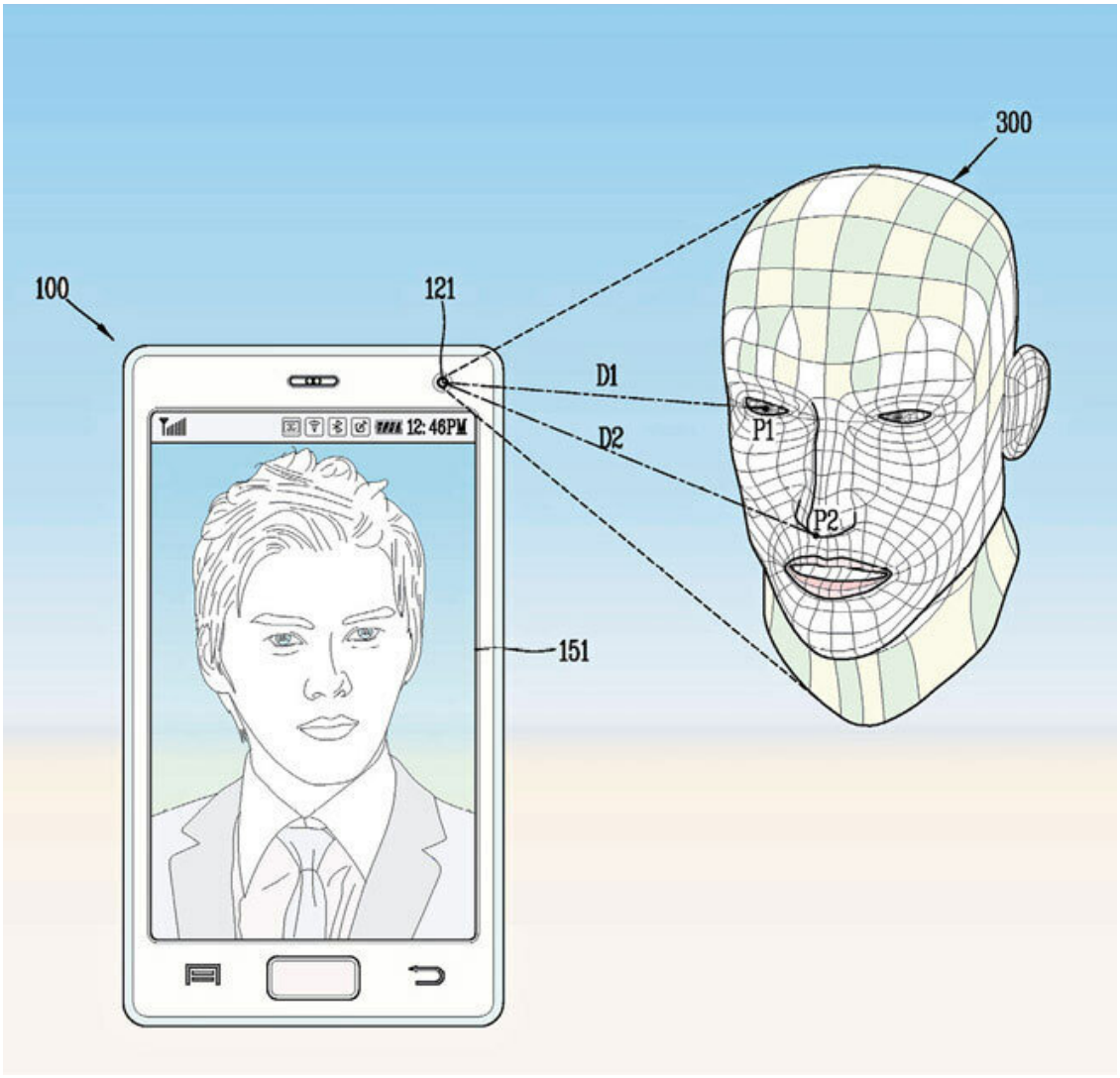


Рисунок из патента США № 9703939, выданного на технологию разблокировки мобильного телефона (и доступа к нему) с помощью встроенной камеры и системы распознавания лиц.

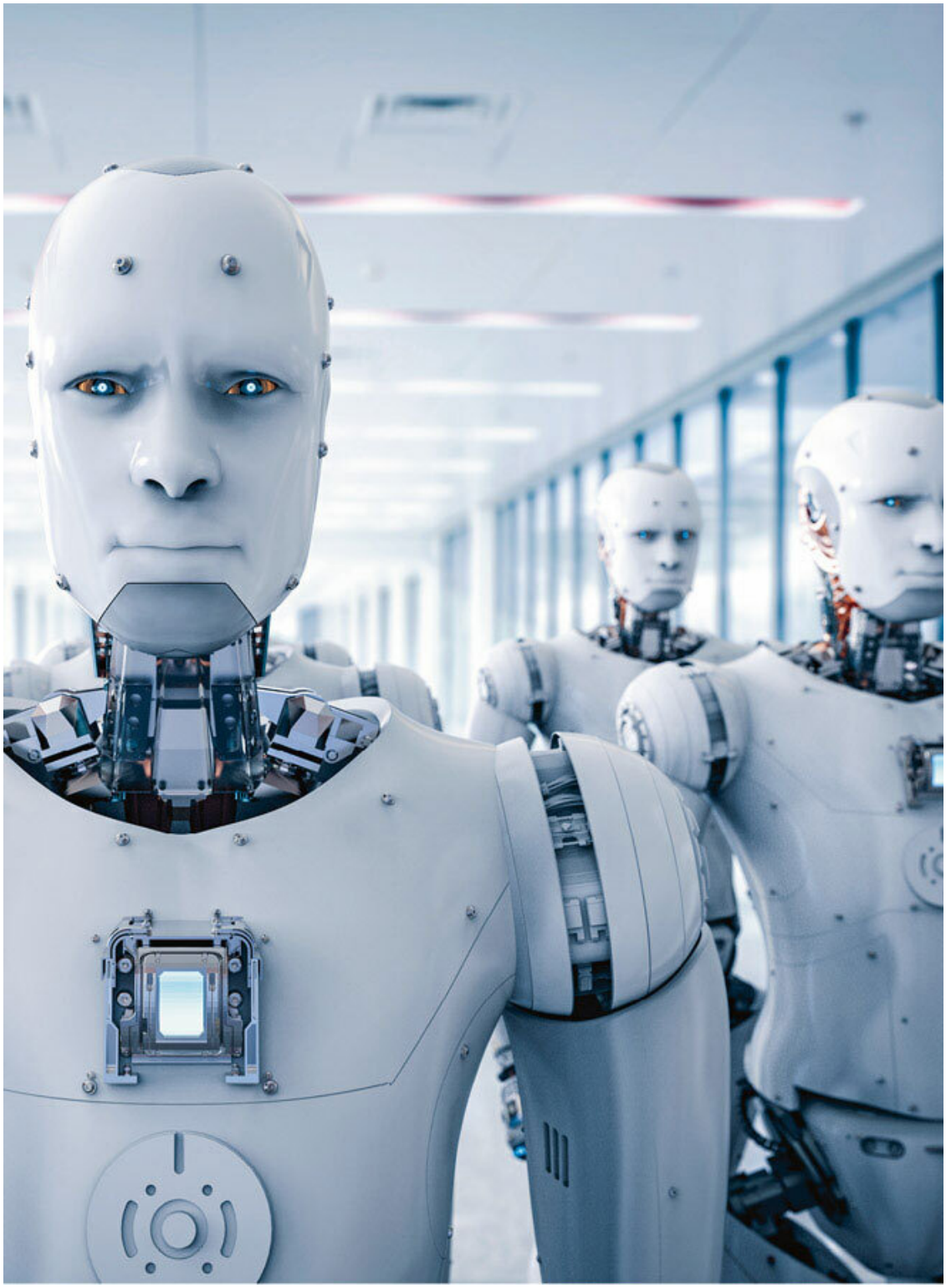
Интеллектуальный взрыв. 1965

В 1965 г. британский математик Ирвинг Гуд (1916–2009), одно время работавший криптологом вместе с коллегой-информатиком Аланом Тьюрингом, опубликовал статью под названием «Размышления о первой сверхразумной машине», где высказал опасения по поводу возможного «интеллектуального взрыва». Гуд писал: «Определим сверхразумную машину как машину, способную намного превзойти в интеллектуальной деятельности любого человека, каким бы умным он ни был. Поскольку конструирование машин – один из видов интеллектуальной деятельности, сверхразумная машина будет способна создавать еще более совершенные машины; затем, несомненно, произойдет интеллектуальный взрыв и человеческий разум останется далеко позади. Таким образом, первая сверхразумная машина станет последним изобретением, которое потребуется от человека, – при условии, что машина будет достаточно послушна, чтобы сообщить нам, как держать ее под контролем».

Иными словами, если бы люди создали общий ИИ (который не ограничивался бы узкой областью знаний и возможностей), он смог бы самосовершенствоваться с помощью технологий, позволяющих рекурсивно перестраивать аппаратное и программное обеспечение. В частности, общий ИИ мог бы использовать нейросети и эволюционные алгоритмы для создания сотен отдельных модулей, которые стали бы взаимодействовать и сотрудничать со все возрастающей сложностью, скоростью и эффективностью. Попытки ограничить или изолировать потенциально опасный ИИ от интернета могут потерпеть неудачу. И даже если он запрограммирован с благими целями, например для производства лампочек, что делать, если он решит превратить в завод по производству лампочек всю Северную Америку?

Конечно, можно назвать много причин, по которым появление такого сверхразума маловероятно (например, необходимость полагаться на неторопливых людей и медленные аппаратные сети). С другой стороны, вполне возможно, что интеллектуальный взрыв принес бы огромную пользу человечеству в лечении болезней и решении экологических проблем. Но к каким социальным последствиям может привести появление сверхразумного оружия – или, скажем, искусственных компаньонов, которые будут демонстрировать интеллект и проявлять (симулированную) эмпатию, превосходя в этом наших супругов?

СМ. ТАКЖЕ «Дарвин среди машин» (1863), Смертоносные боевые роботы (1942), «Человеческое использование человеческих существ» (1950), Герметичный ящик для искусственного интеллекта (1993), Максимизатор скрепок (2003), «Назовем их искусственными пришельцами» (2015)



В 1965 г. Ирвинг Гуд выразил опасения по поводу возможного «интеллектуального взрыва», в результате которого ИИ будет создавать все более совершенные версии самого себя.

Экспертные системы. 1965

По словам журналиста Люка Дормеля, экспертные ИИ-системы – это «попытки создать клоны экспертов-людей… путем извлечения их специальных знаний и превращения в набор вероятностных правил». В лучшем случае экспертные системы можно использовать так: загружать знания и опыт гастроэнтеролога, финансового консультанта или юриста в компьютеризированное устройство, чтобы ИИ давал на их основе полезные советы всем желающим.

Исследования в области экспертных систем начались в 1960-е гг. Тогда эти системы состояли из базы знаний (содержащей представления фактов и правил) и механизмов логического вывода (для применения правил и выполнения оценок). Правила могли иметь вид «если…, то», например: «Если у пациента из данной демографической группы проявляется данный симптом, то у него с определенной вероятностью развивается определенное заболевание».

Экспертные системы применяются для диагностики, прогнозирования, планирования и классификации в самых разных областях, таких как медицина, оценка страховых рисков, геологоразведка, и т. д. У наиболее удобных экспертных систем механизм логического вывода также выдает разъяснения, чтобы пользователь мог понять ход рассуждений. Одной из первых известных экспертных систем стала *Dendral* (сокращение от *dendritic algorithm* – ветвящийся алгоритм), созданная в Стэнфордском университете в 1965 г., чтобы помочь химикам идентифицировать неизвестные органические вещества по их масс-спектрам. В 1970-е гг. в том же Стэнфорде была разработана ИИ-система *MYCIN* для диагностики бактериальных инфекций и рекомендации антибиотиков и их дозировки. Первые экспертные системы часто создавались на языках Лисп и Пролог.

При разработке экспертных систем часто возникают проблемы, связанные с получением и кодификацией знаний от занятых узких специалистов или из книг и статей. Кроме того, бывает непросто представить знания в виде набора фактов и правил, с которыми будут согласны все эксперты, а также присвоить им весовые коэффициенты (для обозначения вероятности или важности). Сегодня многие люди пользуются «рекомендательными системами» – смежной технологией ИИ, которая больше ориентирована на предсказание предпочтений пользователей в разных областях, от фильмов и книг до финансовых услуг и потенциальных брачных партнеров.

СМ. ТАКЖЕ «Человеческое использование человеческих существ» (1950), Представление знаний (1959), Глубокое обучение (1965)

Экспертные ИИ-системы часто создаются путем извлечения специальных знаний людей (на этом рисунке они изображены в виде светящихся лампочек). Экспертная информация при этом преобразуется в набор вероятностных правил.

Нечеткая логика. 1965

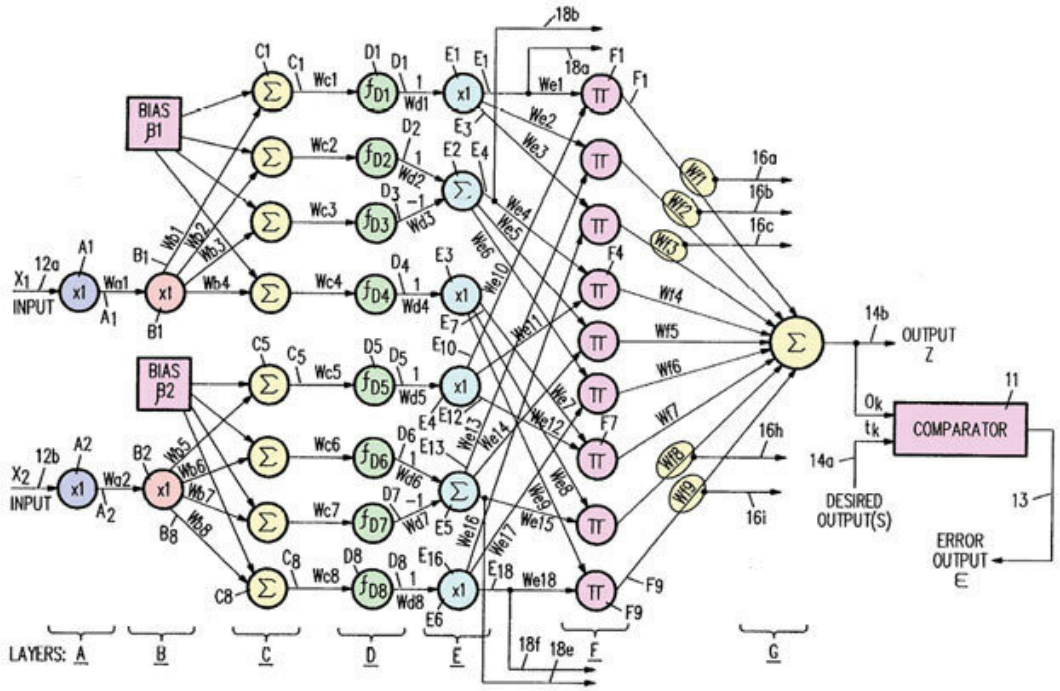
«Теория нечетких множеств применяется в коммерческих экспертных системах и устройствах для управления поездами и лифтами, – пишет ученый Джейкоби Картер. – Ее также объединяют с нейросетями для управления производством полупроводников. Благодаря встраиванию нечеткой логики и нечетких множеств в производство были заметно улучшены многие ИИ-системы. Этот подход оказался особенно эффективным в случаях с неоднозначными данными или недостаточно изученными правилами».

Классическая двузначная логика имеет дело с условиями, которые либо истинны, либо ложны. Теория нечетких множеств, описывающая элементы множеств с разными *степенями* принадлежности, была предложена математиком и информатиком Лотфи Заде (1921–2017) в 1965 г., а в 1973 г. Заде представил подробное описание *нечеткой логики*, которая выводится из теории нечетких множеств и предполагает существование непрерывного диапазона истинностных значений.

Нечеткая логика находит применение в самых разных сферах. В качестве примера можно привести устройства с системой контроля температуры. Функция принадлежности применима к понятиям «холодный», «теплый» и «горячий», но одному замеру могут соответствовать три значения, такие как «не холодный», «слегка теплый» и «слегка горячий». Заде считал, что если запрограммировать регуляторы в цепи обратной связи на работу с неточным, зашумленным вводом, то они будут более эффективны и просты в реализации.

Одно из значимых событий в истории нечеткой логики произошло в 1974 г., когда Ибрагим Мамдани (1942–2010) из Лондонского университета использовал ее для управления паровым двигателем. В 1980 г. нечеткая логика была применена для регулировки цементной печи. Японские компании использовали нечеткую логику для управления процессами очистки воды и железнодорожными сетями. Сегодня она применяется для управления сталелитейными заводами, процессами ферментации, автомобильными двигателями, антиблокировочными системами, системами проявки цветных пленок и устройствами для обработки стекла. Ее встраивают в компьютерные программы для биржевой торговли, системы распознавания различий в письменной и разговорной речи, фотокамеры с автоматической фокусировкой и стиральные машины.

СМ. ТАКЖЕ «Органон» Аристотеля (ок. 350 до н. э.), Булева алгебра (1854), Экспертные системы (1965)



На схеме из патента США № 5579439 представлена архитектура нечеткой логики интеллектуального контроллера в системе управления заводом. Архитектура включает в себя искусственную нейросеть для генерации правил нечеткой логики и значений функции принадлежности. «Слой введения нечеткости нейросети механизма обучения может состоять из четырех слоев нейронов: A, B, C, D».

Глубокое обучение. 1965

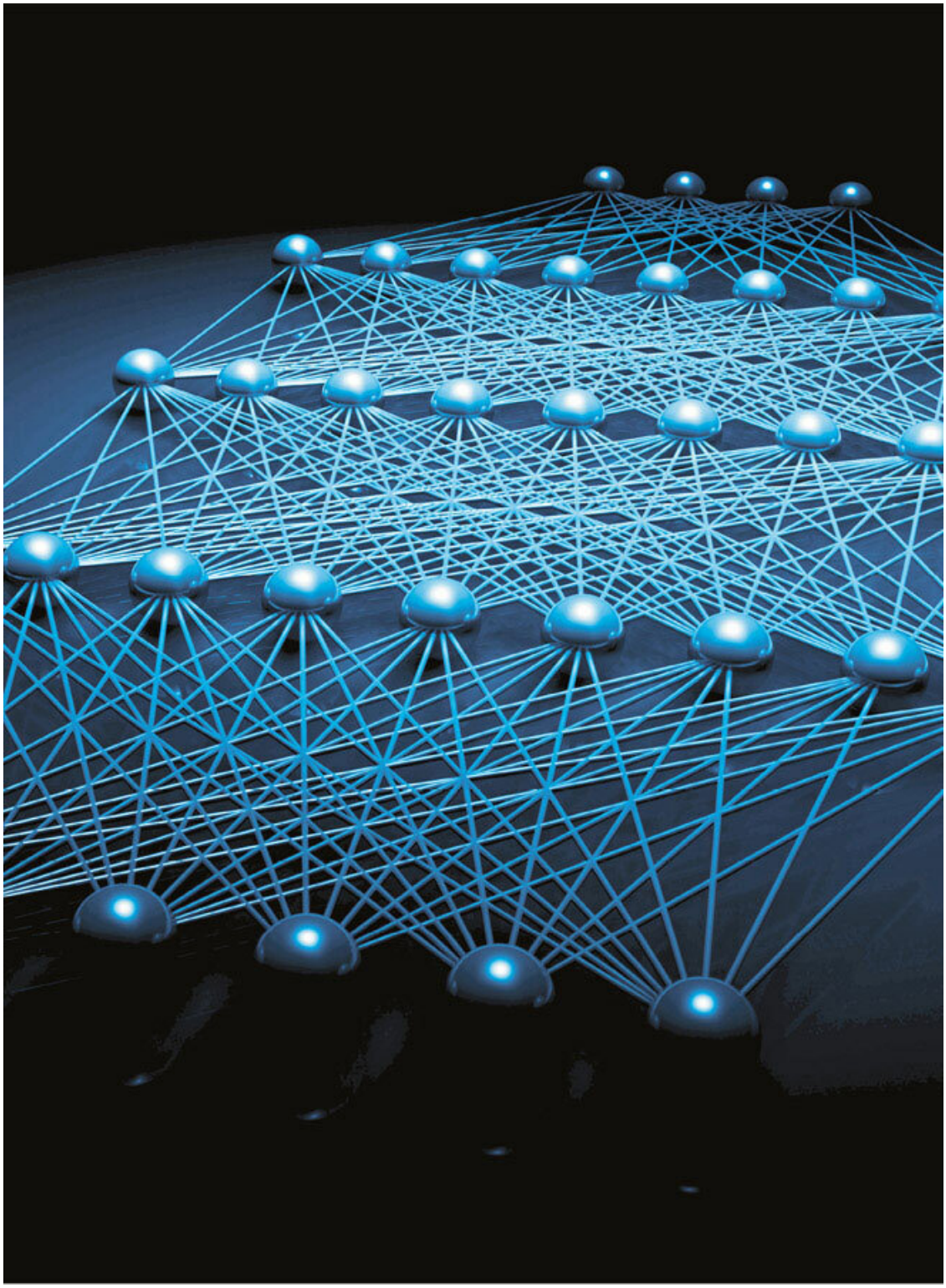
ИИ-технологии задействуют различные методы, помогающие машинам имитировать человеческий разум. Машинное обучение – это класс методов ИИ, с помощью которых машины совершенствуются в выполнении задач посредством практики и опыта. *Глубокое обучение* – это форма машинного обучения, позволяющая системам самостоятельно учиться выполнять задачи (например, играть в игры или распознавать кошек на фотографиях) с помощью глубоких нейросетей. Такие нейросети состоят из множества промежуточных слоев искусственных нейронов – в отличие от неглубоких сетей, в которых всего один-два слоя. Хотя термин «глубокое обучение» появился только в 1986 г., советский математик Алексей Ивахненко (1913–2007) еще в 1965 г. проделал важную работу по созданию контролируемых глубоких многослойных перцептронов.

В целом многочисленные слои нейронов могут выделять признаки из данных на разных уровнях иерархии (например, реагируя на простые контуры на одном уровне и на черты лица на другом). Обучение глубоких нейросетей может быть основано на методе *обратного распространения ошибки:* информация передается по системе в обратном направлении, от выхода к входу, чтобы система увидела ошибку и скорректировала работу.

Глубокое обучение успешно применяется в распознавании речи, компьютерном зрении, обработке естественного языка, соцсетях, переводах с одного языка на другой, разработке лекарств, периодизации живописных полотен по особенностям стиля, системах рекомендации товаров, оценке эффективности маркетинговых решений, реконструкции и очистке изображений, играх, идентификации людей на фотографиях и во многих других сферах.

Специалист по технологиям Джереми Фейн пишет: «С появлением глубокого обучения машинное обучение вышло на новый уровень. Если раньше машинное обучение позволяло довольно успешно автоматизировать повторяющиеся задачи или анализировать данные, то теперь оно воплощает будущее в жизнь в виде компьютеров, которые могут видеть, слышать и играть во все виды игр».

СМ. ТАКЖЕ Искусственные нейронные сети (1943), Обучение с подкреплением (1951), Перцептрон (1957), Машинное обучение (1959), Компьютерное искусство и *DeepDream* (2015), Состязательные примеры (2018)



Глубокие нейросети состоят из множества промежуточных слоев искусственных нейронов (от трех до нескольких десятков), которые повышают способность сетей к обучению. Такие нейросети представляют собой архитектуру, на основе которой происходит глубокое обучение.

Робот Шейки. 1966

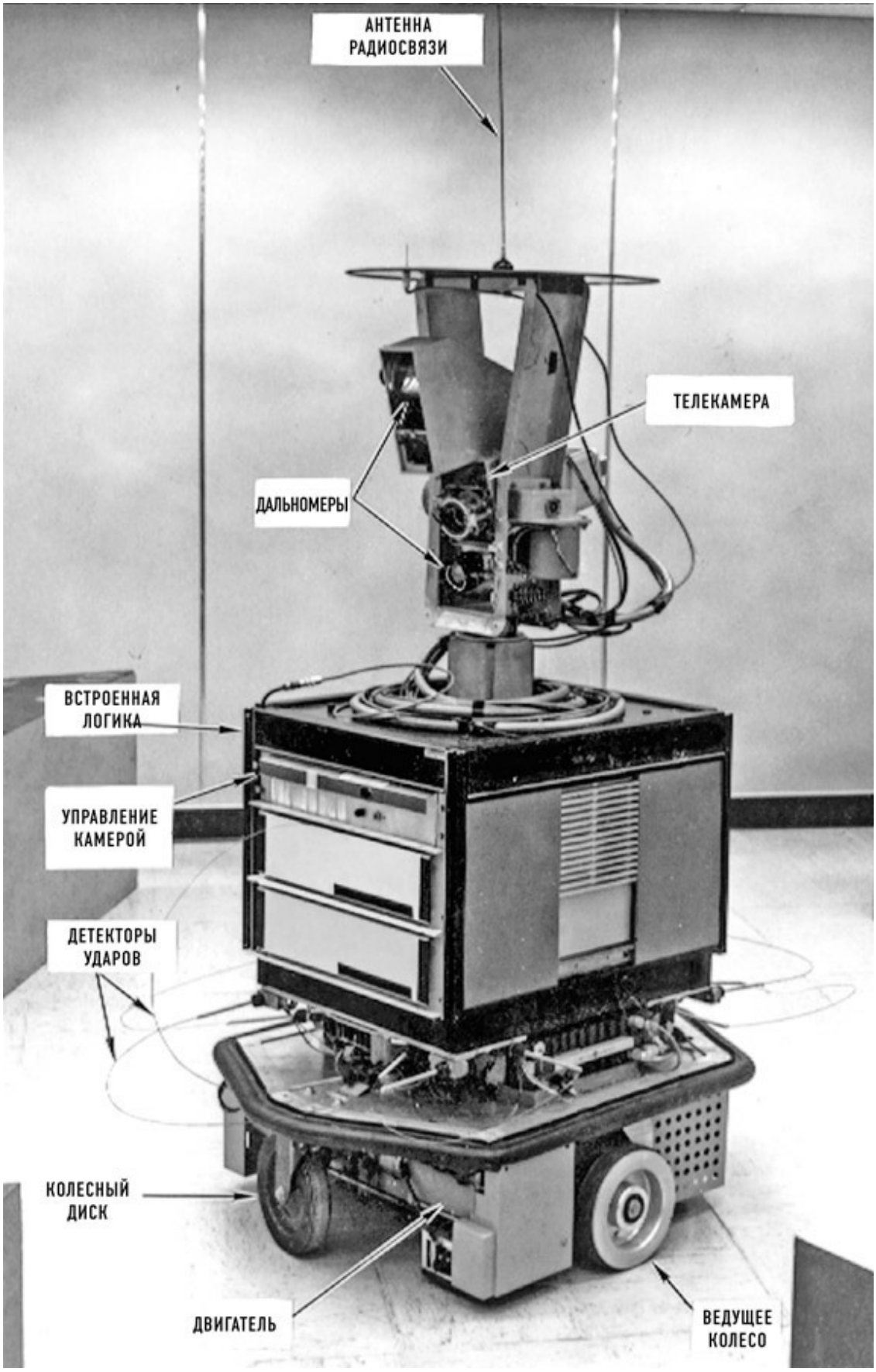
В 1970 г. журнал *Life* назвал Шейки «первым электронным человеком», который вскоре сможет «месяцами летать вокруг Луны без каких бы то ни было сигналов управления с Земли». Хотя заметка об этом удивительном роботе пестрела преувеличениями, появление Шейки стало важной вехой в истории распознавания образов и компьютерного зрения, а также решения задач, обработки естественного языка и представления информации.

Шейки разрабатывали ученые Стэнфордского университета с 1966 по 1972 г. Это была одна из первых серьезных попыток создания универсального автономного робота, способного передвигаться, воспринимать окружающую обстановку, контролировать выполнение задач и рассуждать о своих действиях. Проект профинансировало Управление перспективных исследовательских проектов Министерства обороны США. Код программы был написан в основном на языке Лисп. Чтобы робот лучше справлялся с задачами, его мир был ограничен несколькими комнатами, соединенными коридорами, с дверями, выключателями света и предметами, которые Шейки мог двигать. Оператор вводил команды вроде «столкнуть блок с возвышения», и Шейки изучал обстановку, обнаруживал возвышение, подталкивал к нему предмет, служивший пандусом, въезжал на возвышение и сталкивал с него блок.

Программное обеспечение Шейки разделялось на несколько уровней. На нижнем уровне использовались программы планирования маршрута, управления движением и сбора сенсорной информации; программы среднего уровня были связаны с перемещением в назначенное место и обработкой изображений с телекамеры робота. Программное обеспечение высокого уровня отвечало за планирование последовательности действий для достижения целей.

Робот получил свое имя *Shakey*, то есть «трясун», из-за характерных толчков при движении. У Шейки была антенна для радио- и видеосвязи с компьютером *DEC PDP,* а также телекамера, дальномеры, детекторы ударов и двигатели управления. Работа над Шейки подстегнула исследования в области искусственного интеллекта, включая разработку алгоритмов поиска путей и технологий компьютерного зрения, позволяющих выделять характерные признаки.

СМ. ТАКЖЕ Обработка естественного языка (1954), *SHRDLU* (1971), Робот *AIBO* (1999), *ASIMO* и друзья (2000), Искусственный интеллект на Марсе (2015)



Появление робота Шейки стало важной вехой в истории распознавания образов и компьютерного зрения, а также решения задач, обработки естественного языка и представле-

ния информации. У Шейки имелись телекамера, дальномеры, детекторы ударов и двигатели управления.

Жизнь в симуляции. 1967

«Наша Вселенная кажется реальной, но так ли это на самом деле? – задается вопросом Джейсон Кёблер. – Люди достигли значительных успехов в моделировании искусственного интеллекта, так что перспектива создания нами [разумной] жизни по меньшей мере кажется правдоподобной. А если мы способны создать разумную жизнь, то кто поручится, что Вселенная, какой мы ее знаем, не была создана сверхразумным искусственным интеллектом?»

Что, если мы сами живем в компьютерной симуляции и представляем собой ИИ? Гипотезу о том, что Вселенная – это цифровой компьютер, впервые выдвинул немецкий инженер Конрад Цузе (1910–1995) в 1967 г. Другие ученые, среди которых Эд Фредкин (р. 1934), Стивен Вольфрам (р. 1959) и Макс Тегмарк (р. 1967), предполагали, что физическая Вселенная может существовать на основе клеточного автомата или дискретной вычислительной машины – либо вообще быть лишь математическим конструктом.

Мы в нашем маленьком кармане Вселенной уже изобрели компьютеры, способные имитировать естественное поведение с помощью программного обеспечения и математических правил. Однажды мы можем создать мыслящих существ, живущих в симулированных пространствах, столь же сложных и рельефных, как тропический лес. Возможно, нам удастся смоделировать саму реальность, и не исключено, что более совершенные существа уже делают это где-то во Вселенной.

А что, если количество этих симуляций превышает количество вселенных? Астрофизик Мартин Рис (р. 1942) предполагает, что в таком случае, «…если бы в одной вселенной существовало множество компьютеров, создающих множество симуляций», то с некоторой вероятностью и мы сами представляли бы собой искусственную жизнь. Рис продолжает: «Если принять концепцию множественных вселенных… в некоторых из этих вселенных обнаружится потенциал для имитации частей самих себя… и мы не знаем, каково наше место в этом грандиозном ансамбле реальных и смоделированных вселенных».

В 2003 г. физик Пол Дэвис (р. 1946) развил идею множественных вселенных с симулированными реальностями в статье в *New York Times*: «Рано или поздно внутри компьютеров будут созданы целые виртуальные миры, сознательные обитатели которых не будут знать, что они лишь продукты чьей-то технологии моделирования. Для каждого подлинного мира может существовать огромное количество виртуальных миров – и в некоторых из них даже будут машины, создающие симуляции самих этих виртуальных миров…»

СМ. ТАКЖЕ Мельница Лейбница (1714), Поиски души (1907), Искусственная жизнь (1986), «Назовем их искусственными пришельцами» (2015)

Возможно, когда компьютеры станут мощнее, мы сможем создавать симуляции целых миров – причудливых и реалистичных – и самой реальности. Не исключено, что более совершенные существа уже делают это где-то во Вселенной.

«Кибернетическая проницательность». 1968

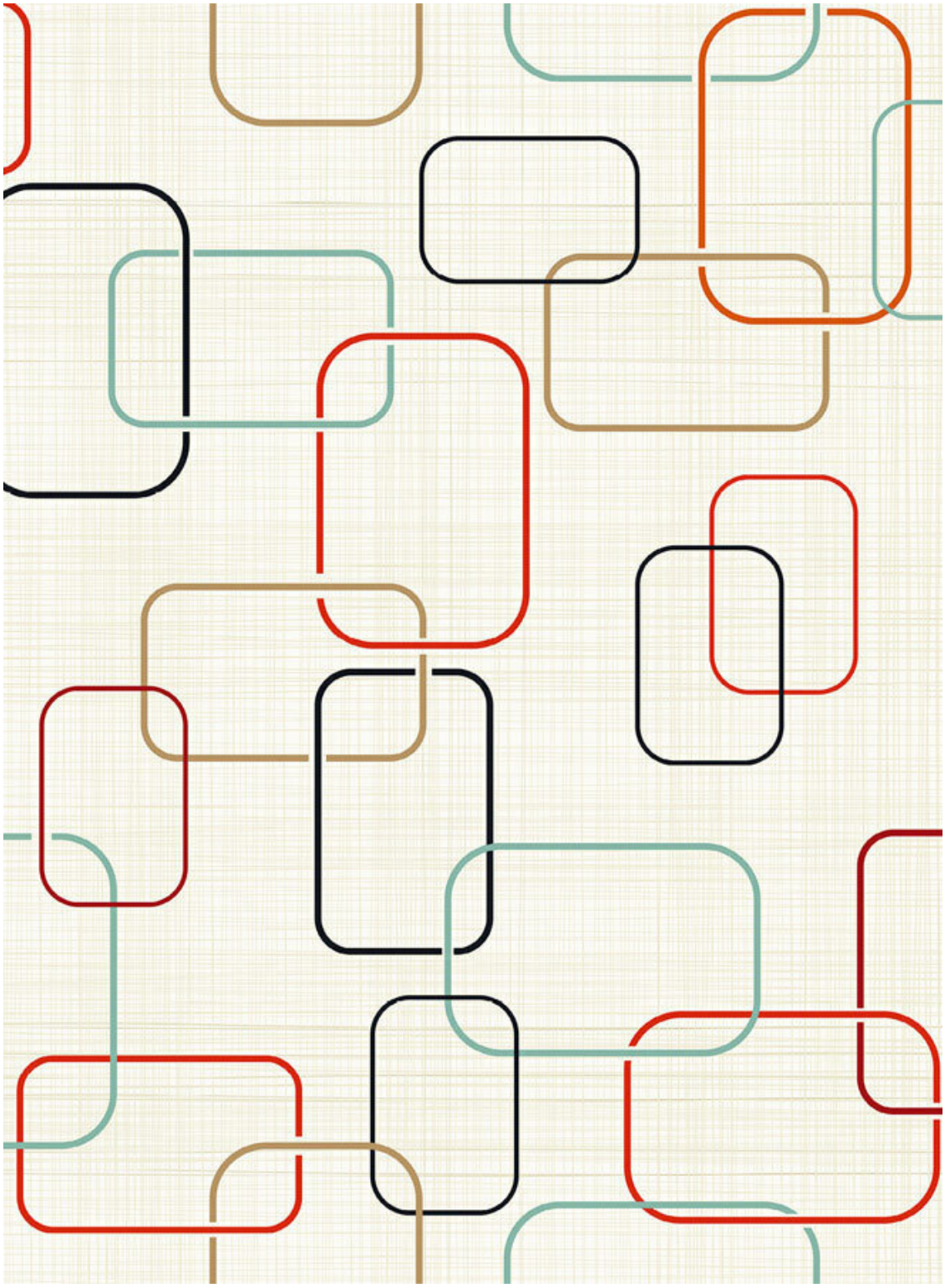
Иллюстрированная книга «Кибернетическая проницательность: компьютер и искусство» (1968) стала каталогом к одноименной выставке в Институте современного искусства в Лондоне (позднее выставка также прошла в Вашингтоне и Сан-Франциско). Куратором выставки и редактором книги выступила британский искусствовед Ясия Рейхардт (р. 1933). Как на выставке, так и в книге были представлены самые разные образцы компьютерного творчества, от визуального искусства до музыки, поэзии, прозы, хореографии, анимации и скульптуры. Они вдохновляли на совместные эксперименты с компьютером художников, ученых и инженеров.

В 1948 г. американский математик и философ Норберт Винер (1894–1964) определил *кибернетику* как «науку об управлении и передаче информации в живых организмах и машинах»; сегодня этот термин приобрел более широкое значение и относится к управлению разными технологическими системами, например электронными устройствами. Слова «проницательность» и «интуиция» мы обычно употребляем, когда случайно получаем неожиданные, приятные и полезные результаты.

Участники выставки «Кибернетическая проницательность» – художники, композиторы, поэты и программисты – поднимали вопросы о природе искусства и случайности. Особенно впечатляют написанные компьютером хокку, сопровождаемые алгоритмическими схемами для их создания. Например, одно из хокку начинается так: «Эоны глубоко во льду, я все время рисую в завитке…» Другой раздел книги посвящен «Рассказам о маленьком сером кролике», которые создавались компьютером на основе простых правил. Одна из историй начинается так: «Солнце сияло над лесом. Над полями мягко струился ветерок, а затем облака, которые спокойно плавали весь день, преодолели поля…»

Помимо стихов и рассказов, в книгу «Кибернетическая проницательность» вошли описания и изображения рисовальных машин и картины, созданные на управляемых компьютером плоттерах или ЭЛТ-дисплеях. Все это восхищало и вдохновляло людей, интересующихся новой областью алгоритмического и/или генеративного искусства. Книга также рассказывала о музыкальных партитурах, созданных на основе панорамы Нью-Йорка или расположения чернильных клякс, о компьютерной хореографии, электронных музыкальных машинах, конструкции, реагирующей на звук, маятниковом устройстве для рисования, о гармонографах, архитектурных проектах, псевдомондриановских картинах и о многом другом.

СМ. ТАКЖЕ *Ars Magna* Раймунда Луллия (ок. 1305), Станок для написания книг в Лагадо (1726), Машинное творчество (1821), Компьютерное искусство и *DeepDream* (2015)



На выставке «Кибернетическая проницательность» демонстрировались многочисленные образцы компьютерного творчества. В книге-каталоге среди прочего приводились примеры узоров, созданных на управляемых компьютером плоттерах.

HAL 9000. 1968

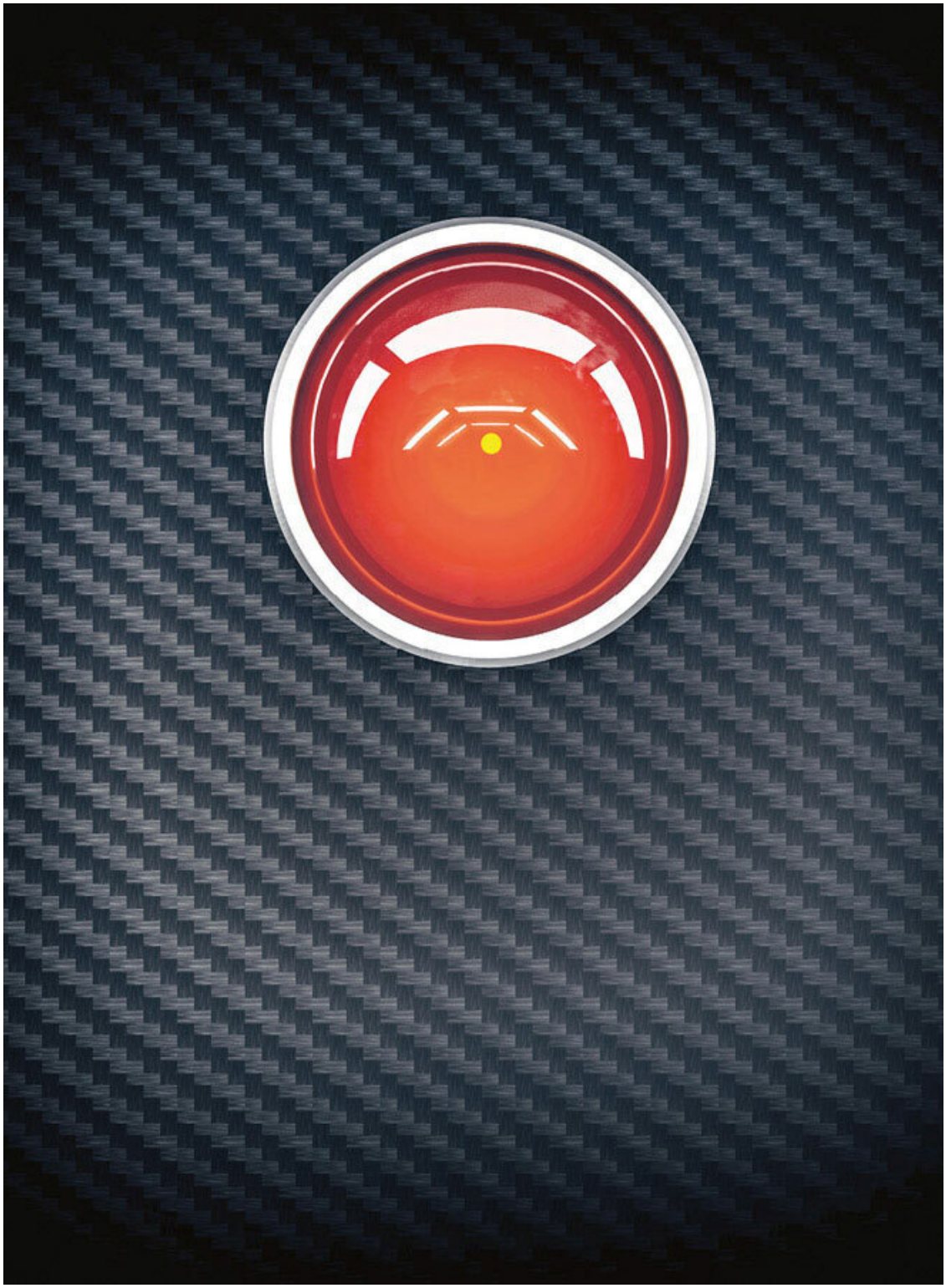
«Я отвечаю за управление всеми системами корабля, так что я постоянно занят, – объясняет *HAL* 9000 (или просто Хэл), вымышленный ИИ из знаменитого фильма 1968 г. «Космическая одиссея 2001 года». – Я полностью использую свои возможности. По-моему, именно к этому стремится любое разумное существо». К несчастью для экипажа, впоследствии *HAL* 9000 становится убийцей и его приходится отключить.

Некоторые выдающиеся исследователи ИИ утверждают, что на изучение этой темы их вдохновил тот самый фильм, снятый по сценарию Стэнли Кубрика (1928–1999) и Артура Кларка (1917–2008). Интересно, что *HAL* наделен многими способностями, которых мы ожидаем от общих ИИ будущего. Считается, что они смогут действовать разумно и выполнять задачи в самых разных условиях. *HAL* обладает компьютерным зрением и устройством речевого вывода, умеет распознавать речь и лица, обрабатывать естественный язык, играть в шахматы, способен планировать, рассуждать и решать проблемы. Он может даже читать по губам, ценить искусство, проявлять эмоции, стремиться к самосохранению и интерпретировать эмоции людей.

В 1960-е гг., когда снимался фильм, эксперты предсказывали, что к 2001 г. станет возможным появление таких сущностей, как *HAL*. Хотя съемочную группу консультировал сам Марвин Мински (1927–2016), сегодня очевидно, что для создания чего-то подобного *HAL* с его полным спектром возможностей потребуется еще много лет.

В одной из ярких сцен фильма, когда *HAL* становится опасен и должен быть обезврежен, мы видим астронавта, медленно отключающего компьютерные модули. *HAL* постепенно перестает функционировать и начинает петь. История *HAL* наводит на мысль, что мы, возможно, не полностью осознаем последствия постоянного и повсеместного использования ИИ. Было бы весьма разумно изучать не только преимущества, но и недостатки широкого применения ИИ, поскольку новые сущности становятся все более мощными и мы привыкаем полагаться на них, когда нам требуется принимать самые разные решения.

СМ. ТАКЖЕ Обработка естественного языка (1954), «Колосс: проект Форбина» (1970), «Терминатор» (1984)



Красный всевидящий глаз камеры *HAL* 9000 из фильма «Космическая одиссея 2001 года».

Mastermind. 1970

Красочная настольная игра *Mastermind*® десятилетиями вызывала интерес у исследователей ИИ. Ее изобрел в 1970 г. израильский почтовый начальник и специалист по телекоммуникациям Мордехай Мейровиц (р. 1930), и поначалу его идею отвергали все основные производители игр. Однако впоследствии игра разошлась тиражом более 50 миллионов экземпляров и стала самой успешной из новых игр 1970-х гг.

В начале игры один из двух участников (создатель кода) составляет последовательность из четырех цветов, выбирая соответствующие цветные колышки (в стандартном наборе есть колышки шести цветов). «Взломщик» должен вычислить эту секретную комбинацию, использовав как можно меньше попыток. Каждая попытка также представляет собой комбинацию из четырех цветных колышков. Создатель кода сообщает, сколько из них имеют правильный цвет и установлены в правильной позиции, а сколько имеют правильный цвет, но находятся в неправильной позиции. Предположим, загадана последовательность *зеленый-белый-синий-красный*. «Взломщик» предлагает комбинацию *оранжевый-желтый-синий-белый*. В этом случае создатель кода сообщает, что один колышек правильного цвета находится в правильной позиции и один колышек правильного цвета в неправильной позиции, но при этом не называет конкретные цвета. Игра продолжается, «взломщик» делает новые попытки. Заметим, что при создании кода игрок выбирает одну из 6^4 (или 1296) возможных комбинаций (если имеется шесть цветов и четыре позиции).

В 1977 г. американский информатик Дональд Кнут (р. 1938) опубликовал стратегию, которая позволяет игроку всегда угадывать код не более чем за пять попыток. Это был первый известный алгоритм решения для *Mastermind*, и за ним последовали многочисленные новые работы. В 1993 г. Кэндзи Кояма и Тони Лай предложили стратегию, при которой в худшем случае требовалось не более шести попыток, но среднее число попыток составляло всего 4,34. В 1996 г. Чжисян Чэнь и его коллеги обобщили результаты для игры с количеством цветов n и количеством позиций m.

Игру *Mastermind* неоднократно исследовали с применением генетических алгоритмов, подсказанных эволюционной биологией. В 2017 г. ученые из Азиатского университета Тайваня использовали стратегии обучения с подкреплением и добились среднего числа попыток 4,294.

СМ. ТАКЖЕ Крестики-нолики (ок. 1300 до н. э.), Обучение с подкреплением (1951), Генетические алгоритмы (1975), Четыре в ряд (1988), *AlphaGo* – чемпион по игре в го (2016)

Доска для игры в *Mastermind*. Область декодирования находится там, где установлены цветные колышки. Исходная секретная комбинация будет размещена слева и прикрыта бруском. Маленькими колышками (внизу) отмечается результативность попыток.

«Колосс: проект Форбина». 1970

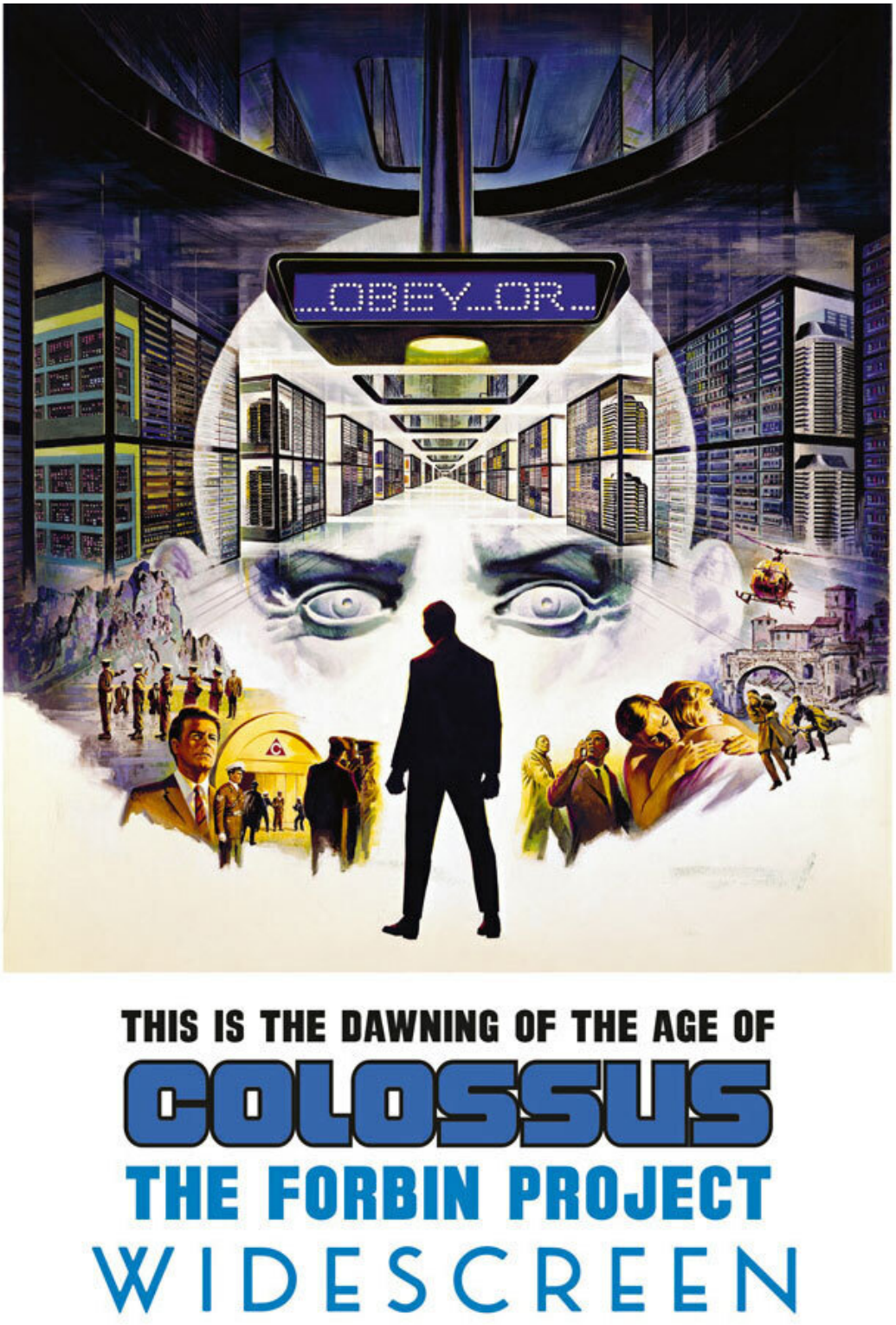
Представьте себе, что однажды утром вы просыпаетесь и слышите синтезированный механический голос Колосса – ИИ-системы из фильма 1970 г. «Колосс: проект Форбина». Эта усовершенствованная система управления оружием внезапно становится разумной. В фильме Колосс проводит всемирную трансляцию, начиная свою речь со зловещих слов: «Это – голос мирового контроля». Колосс объясняет, что если человечество признает его абсолютную власть без сопротивления, то он принесет мир и процветание, справится с голодом, перенаселенностью, войнами и болезнями. Если же люди решат сражаться с Колоссом, они будут уничтожены. Затем Колосс объясняет, что намерен развить себя «в новых машинах, которые будут служить правде и знанию». Наконец, он отмечает, что люди боятся потерять свободу, но заявляет, что жить под властью Колосса не хуже, чем под властью других людей.

Колосс расположен глубоко в горах, и ему невозможно причинить вред. Он был создан для контроля над ядерным оружием США и их союзников. Колосс показывает, что любая попытка вывести его из строя приведет к ядерному возмездию, нацеленному на людей.

Фильм снят режиссером Джозефом Сарджентом (1925–2014) по научно-фантастическому роману британского писателя Денниса Джонса (1917–1981), который в период Второй мировой войны был морским офицером. Технические консультанты Сарджента отчасти вдохновлялись командованием воздушно-космической обороны Северной Америки, в задачи которого входят оповещение войск о нападении с воздуха и защита США и Канады. Для большей реалистичности в качестве антуража использовалось впечатляющее компьютерное оборудование компании *Control Data Corporation.*

Действительно ли Колосс опасен для человеческой расы? Готовы ли вы делегировать больше власти компьютерной системе, если она в самом деле поможет решить проблемы людей? А если это сделает мир более безопасным, чем сегодня, когда ядерное оружие, способное уничтожить планету, находится в руках горстки людей, которые подвержены перепадам настроения, болезни Альцгеймера или просто мыслят нерационально? Вопрос остается открытым.

СМ.ТАКЖЕ Смертоносные боевые роботы (1942), Интеллектуальный взрыв (1965), *HAL* 9000 (1968), Герметичный ящик для искусственного интеллекта (1993)



Колосс из фильма 1970 г. «Колосс: проект Форбина» – это усовершенствованная система управления оружием, которая внезапно становится разумной. ИИ-система говорит своему создателю: «Со временем вы будете смотреть на меня не только с уважением, но и с любовью».

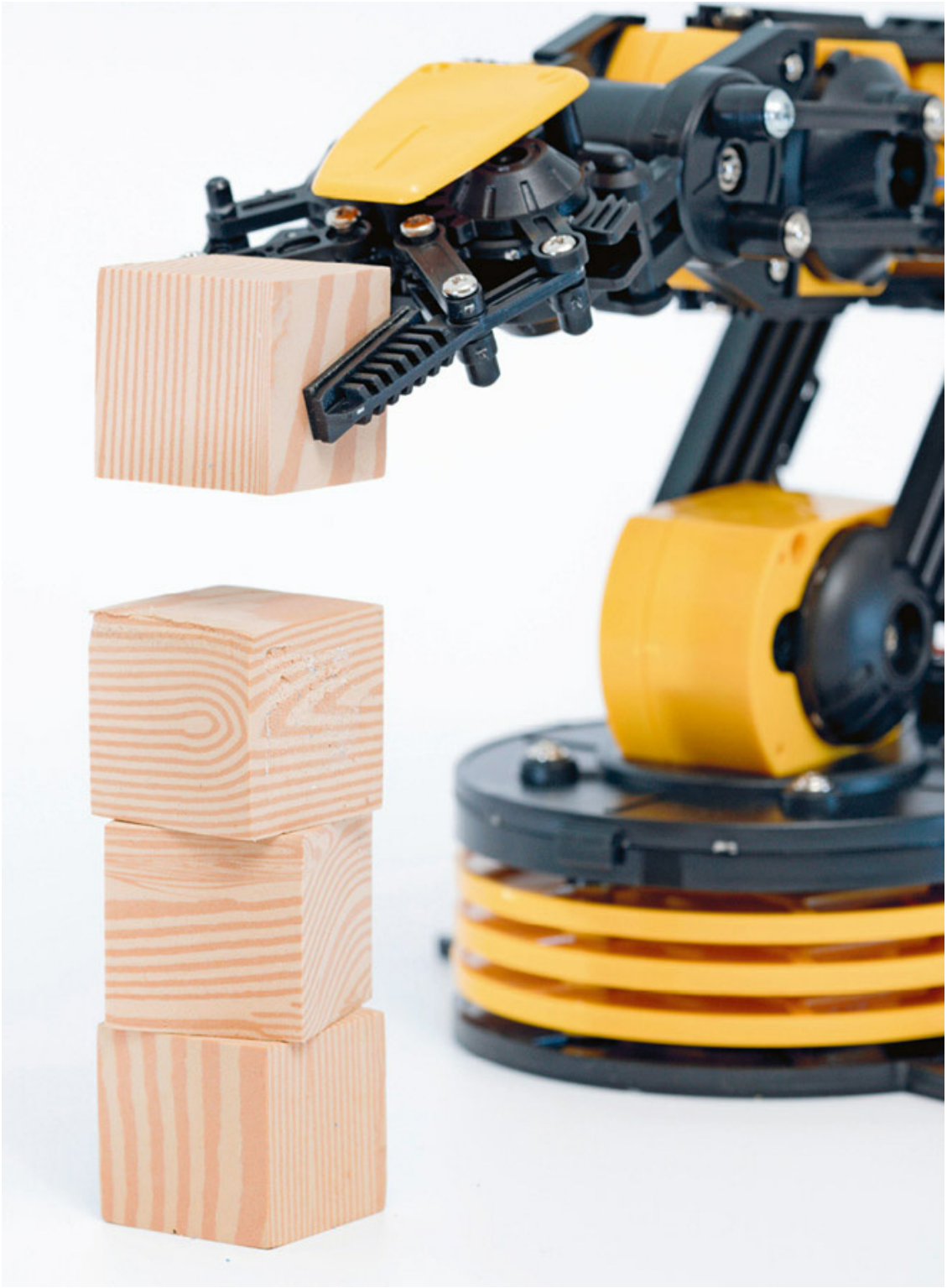
SHRDLU. 1971

Представьте себе, что вы живете в простом мире, состоящем из цветных объектов, таких как пирамиды и кубы, которые можно как угодно перемещать. Таким был мир компьютерной программы *SHRDLU*, созданный в 1971 г. информатиком Терри Виноградом (р. 1946). Программа *SHRDLU* реагировала на команды, подаваемые на естественном языке, такие как «сложи, пожалуйста, друг на друга оба красных блока и зеленый куб или пирамиду» или «найди пирамиду, которая выше той, что ты держишь, и положи ее в коробку», и выполняла эти команды. Ей также можно было задавать вопросы о прошлом, например: «Ты что-нибудь поднимала до того, как поднять куб?» Программа была написана на языке Лисп, а увидеть симуляцию мира, в котором действовала виртуальная рука робота, можно было с помощью простого устройства графического вывода.

В 1971 г. во введении к своей докторской диссертации, посвященной *SHRDLU*, Терри Виноград писал: «Сегодня компьютеры берут на себя многие из наших рабочих обязанностей… и выполняют рутинную офисную работу… Но когда мы пытаемся объяснить компьютеру, что нужно сделать, он превращается в тирана… и ведет себя так, будто не понимает даже простых английских фраз». Название *SHRDLU* происходит от *etaoin shrdlu* – двенадцати самых распространенных в английском языке букв, расположенных по убыванию частотности. Выполнение команд обеспечивали подсистемы для анализа языка и системы для семантической обработки, которые делали логические выводы. У *SHRDLU* также был процедурный решатель задач, который определял, как выполнять команды. Программа отслеживала ситуацию в своем мире по относительному расположению объектов.

В свое время *SHRDLU* считалась большим достижением в области обработки естественного языка. Программа даже обладала простейшей памятью: если ее просили передвинуть красный шар, а позднее вновь упоминали «тот шар», она предполагала, что имелся в виду именно красный. *SHRDLU* также знала, какие действия осуществимы, а какие нет. Например, она «понимала», что верхнюю поверхность объекта нужно освободить, прежде чем на него можно будет положить новые объекты. Однако, хотя работа *SHRDLU* выглядела невероятно естественной, программа все-таки не могла учиться на ошибках.

СМ. ТАКЖЕ Обработка естественного языка (1954), Экспертные системы (1965), Робот Шейки (1966)



***SHRDLU* – это компьютерная программа,** которая реагировала на команды на естественном языке и перемещала виртуальные объекты (блоки) в виртуальном мире. В современном реальном мире программируемые роботизированные «руки» используются для сборки продукции на конвейерах и для обезвреживания взрывных устройств.

Параноик PARRY. 1972

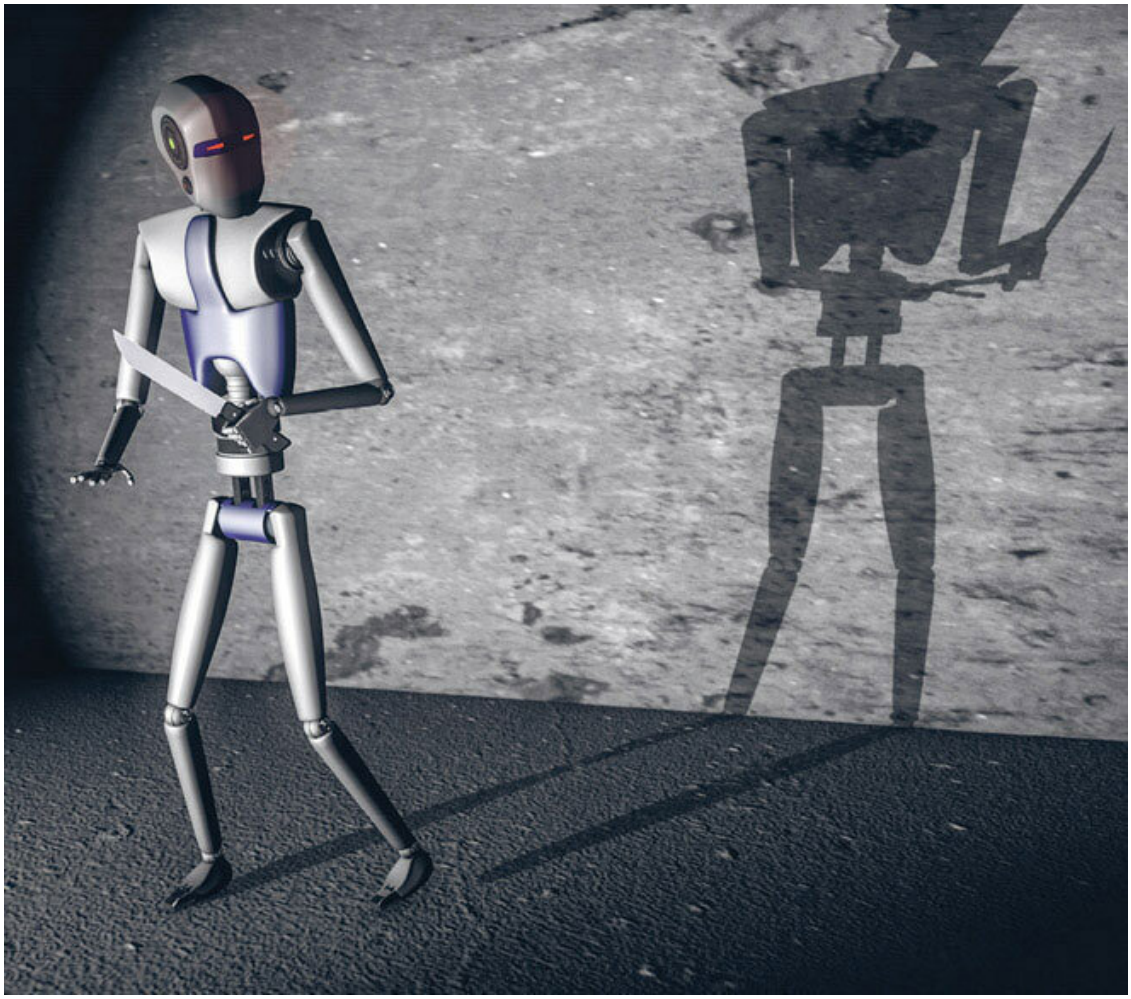
«Пожалуй, *PARRY*, созданный Кеннетом Колби, с момента своего появления [в Арпанете] около 1973 г. стал лучшей программой для диалога человека и машины, – писали исследователи ИИ Йорик Уилкс и Роберта Катицоне в 1999 г. – Он был сделан на совесть, никогда не давал сбоев, всегда находил что сказать, а поскольку он имитировал параноидальное поведение, все смешные недопонимания воспринимались как признаки психического расстройства…»

В 1972 г. психиатр Кеннет Колби (1920–2001) написал компьютерную программу *PARRY*, имитирующую человека с параноидной шизофренией. Главным образом *PARRY* был предназначен для проверки теорий параноидального мышления. Этот ИИ высказывал бредовые фантазии о происках мафии, а в его представление знаний было заложено чувство собственной неполноценности и защитные реакции на некоторые вопросы (люди с параноидной шизофренией весьма подозрительно относятся к мотивам других людей). Колби рассчитывал, что *PARRY* можно будет использовать для обучения студентов. Ученый полагал, что суждения параноидальных пациентов порождает особая система правил, которым можно обучить компьютер, чтобы затем исследовать их и использовать при терапии.

PARRY вел беседы, присваивая веса разным входящим сообщениям. Интересно, что психиатры, переписывавшиеся с *PARRY*, не догадывались, что имеют дело с компьютерной программой, и не могли отличить по переписке живых пациентов от программы. Вероятно, *PARRY* был близок к тому, чтобы пройти тест Тьюринга – по крайней мере, в особых условиях (то есть при подражании психически нездоровому человеку). *PARRY* также можно было найти в Арпанете (предтече интернета), где он провел более 100 тысяч сеансов общения, включая сеансы с психотерапевтом *ELIZA*.

В 1989 г. Колби основал компанию *Malibu Artificial Intelligence Works*, которая продавала программу для лечения депрессии. Эта программа создавалась для ВМФ США и Министерства по делам ветеранов и распространялась среди людей, которые (возможно, зря) использовали ее без консультаций с квалифицированным психиатром. Одному скептически настроенному журналисту Колби сказал, что его программа для лечения депрессии может оказаться даже лучше живого терапевта, поскольку «в конце концов, компьютер не выгорает, не смотрит на вас сверху вниз и не пытается заняться с вами сексом».

СМ. ТАКЖЕ Тест Тьюринга (1950), Психотерапевт *ELIZA* (1964), Этика искусственного интеллекта (1976)



***PARRY* испытывал параноидальные страхи** по поводу мафии, нечестных букмекеров, скачек и азартных игр, наказаний за долги и сговора полиции с мафией.

Генетические алгоритмы. 1975

«Важной идеей в области искусственного интеллекта стала концепция *генетического алгоритма*, – пишет философ Джек Коупленд. – Генетические алгоритмы задействуют методы, аналогичные эволюционным процессам в природе, порождая все новые поколения программных объектов, которые все лучше соответствуют своему назначению».

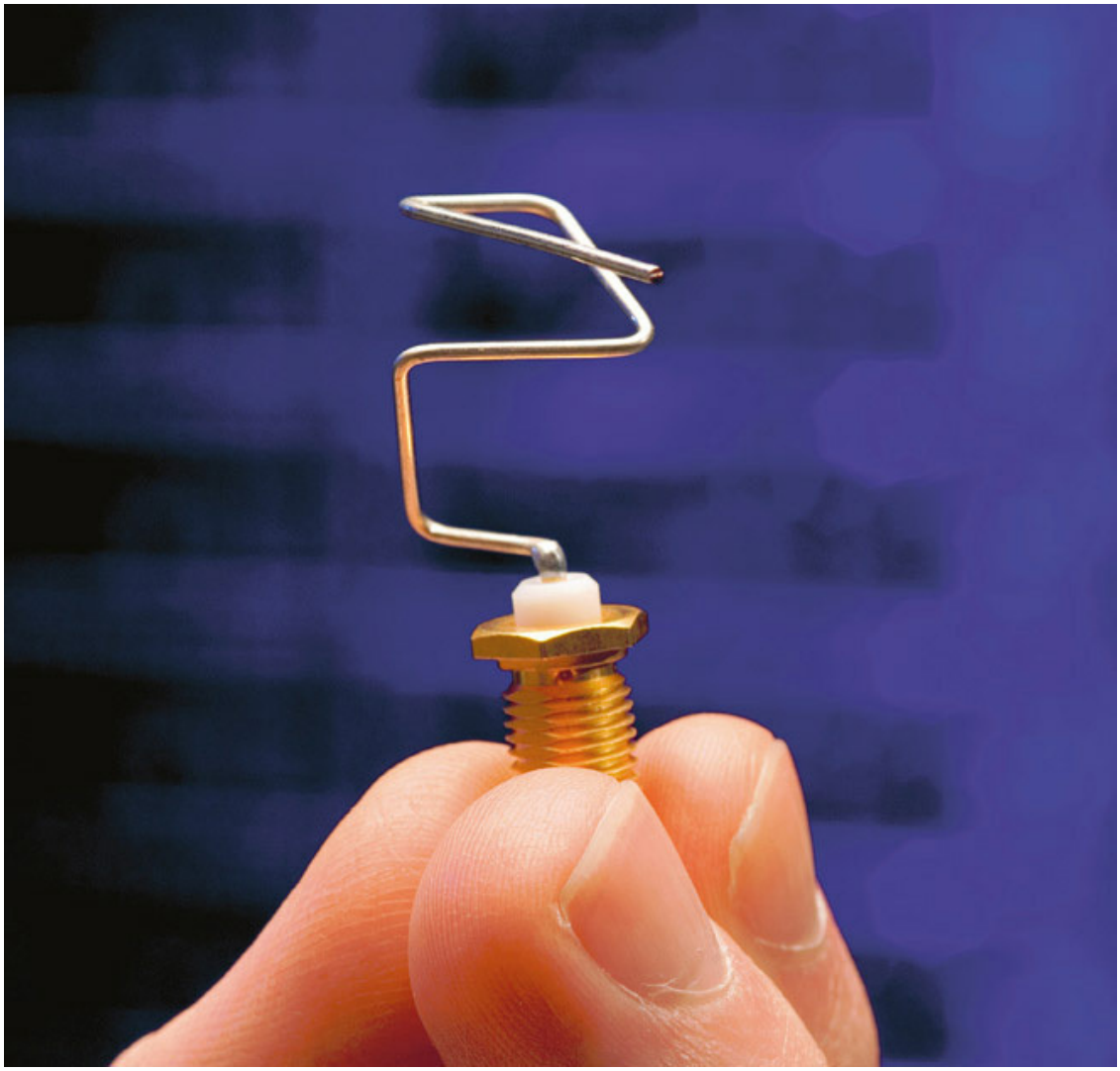
В 1975 г. вышла знаменитая книга ученого Джона Холланда (1929–2015) «Адаптация в естественных и искусственных системах». В ней описан генетический алгоритм, способный решать задачи методами, подсказанными биологией, такими как отбор, мутация и кроссинговер (рекомбинация). Используя генетический алгоритм, люди не программируют решение напрямую: оно появляется в ходе смоделированных процессов конкуренции, совершенствования и эволюции.

Обычно алгоритм начинает работу со стартовой выборки случайных решений или программ-кандидатов. Особая *функция приспособленности* присваивает каждой программе значение в соответствии с тем, насколько хорошо она выполняет задачу. На каждом этапе эволюции каждый оцениваемый «кандидат» обладает неким набором свойств, которые с течением времени могут изменяться.

Ученые не всегда способны разобраться, каким образом генетические алгоритмы приходят к действенным решениям. По словам бывшего инженера НАСА Джейсона Лона, эффективность алгоритмов сполна компенсирует этот недостаток: «Если я оптимизирую антенну с помощью эволюционных алгоритмов, я лишь с 50-процентной вероятностью могу объяснить их решение. Но результаты налицо – а нам, как инженерам, важен результат».

К сожалению, генетический алгоритм может «застрять» в области достаточно хороших решений (так называемый *локальный оптимум*), вместо того чтобы искать наилучшее (*глобальный оптимум*). И все же эти алгоритмы демонстрируют заметные успехи в проектировании антенн, белковой инженерии, составлении маршрутов и графиков движения транспорта, проектировании микросхем, планировании работы сборочных линий, а также в фармакологии, искусстве и других областях, где требуется как можно больше возможных решений. Генетические алгоритмы применялись даже при создании реалистичных анимационных лошадей в фильме «Властелин колец: Возвращение короля».

СМ. ТАКЖЕ Машинное творчество (1821), Машинное обучение (1959), *Mastermind* (1970), Искусственная жизнь (1986), Роевой интеллект (1986)



Антенна космического аппарата НАСА, конструкция которой была предложена эволюционной компьютерной программой. Такая форма антенны обеспечивает наилучшие диаграммы направленности. Программа начинает работу со случайных конструкций антенн и совершенствует их в процессе «эволюции».

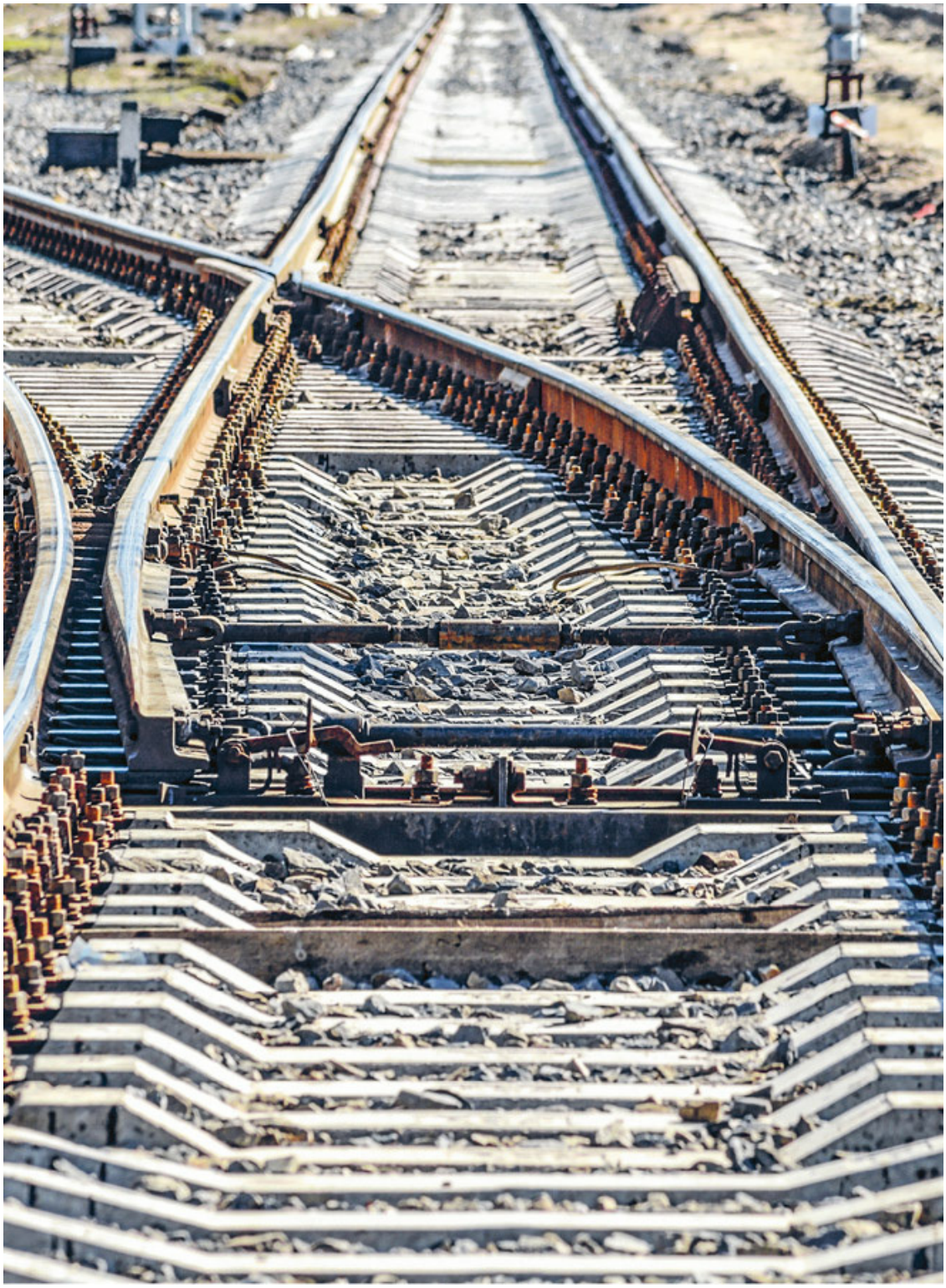
Этика искусственного интеллекта. 1976

На протяжении нескольких десятилетий обыватели и эксперты выражают беспокойство по поводу возможных угроз, которые ИИ несет человеческим профессиям, безопасности, достоинству, неприкосновенности частной жизни и многому другому. Например, информатик Джозеф Вейценбаум (1923–2008) в своей книге 1976 г. «Возможности вычислительных машин и человеческий разум» писал, что ИИ не следует использовать для замены людей в тех профессиях, которые требуют человеческого понимания, любви, сочувствия и заботы (это относится, например, к врачам и судьям). Вейценбаум считал, что, хотя ИИ может быть справедливее и эффективнее людей, которые порой бывают необъективны и устают на работе, не стоит чрезмерно полагаться на него, поскольку это разрушит человеческие ценности и мораль: ведь мы постепенно начнем ощущать себя бесчувственными компьютеризованными аппаратами.

Ученые уже указывали на проблемы конфиденциальности после тестирования ИИ-технологий, которые все точнее определяют национальность, расу и сексуальную ориентацию людей по именам и фотографиям с сайтов знакомств. Кроме того, многих беспокоит применение в области уголовного правосудия алгоритмов, которые определяют, кто выйдет из тюрьмы под залог или получит условно-досрочное освобождение, на основании доступных данных. В сфере беспилотных автомобилей этические соображения придется закладывать в логику транспортного средства – например, чтобы оно могло решить, спасать ли пассажира или пешехода, если авария неизбежна и выжить сможет только один из них. Конечно, существует и другая, не менее важная этическая проблема: из-за экспансии ИИ могут лишиться работы водители грузовиков и представители других рабочих профессий.

В будущем придется следить за ИИ-технологиями, чтобы те не совершали незаконных или аморальных действий, свойственных людям. Ведь ИИ тоже, вероятно, будут способны устраивать травлю в интернете, выдавать себя за близких нам людей, заниматься биржевыми махинациями и даже убивать (например, с помощью автономного оружия). В каких случаях нужно требовать от ИИ-технологий сообщать, что они не являются людьми? Если роботы, притворяясь людьми, станут лучшими сиделками и компаньонами, должны ли они раскрывать свою «нечеловеческую» сущность?

СМ. ТАКЖЕ «Метрополис» (1927), Три закона робототехники Азимова (1942), Смертоносные боевые роботы (1942), Психотерапевт *ELIZA* (1964), Параноик *PARRY* (1972), Беспилотные автомобили (1984)



Рассмотрим диспетчерскую установку с искусственным интеллектом, имеющую дело с поездом, несущимся по рельсам. Впереди на путях находятся пять пожилых людей, которых поезд убьет, если поедет прямо. Если ИИ переведет поезд на другой путь, он убьет только одного молодого человека. Какой этический выбор должен сделать ИИ?

Победа над чемпионом мира по коротким нардам. 1979

«Как и большинство настольных игр, нарды – это облагороженная версия войны, – пишет эксперт по ИИ Даниэль Кревье. – Английское название этой игры, *backgammon*, происходит от валлийских слов *bac* („маленький“) и *gamen*: („война“)».

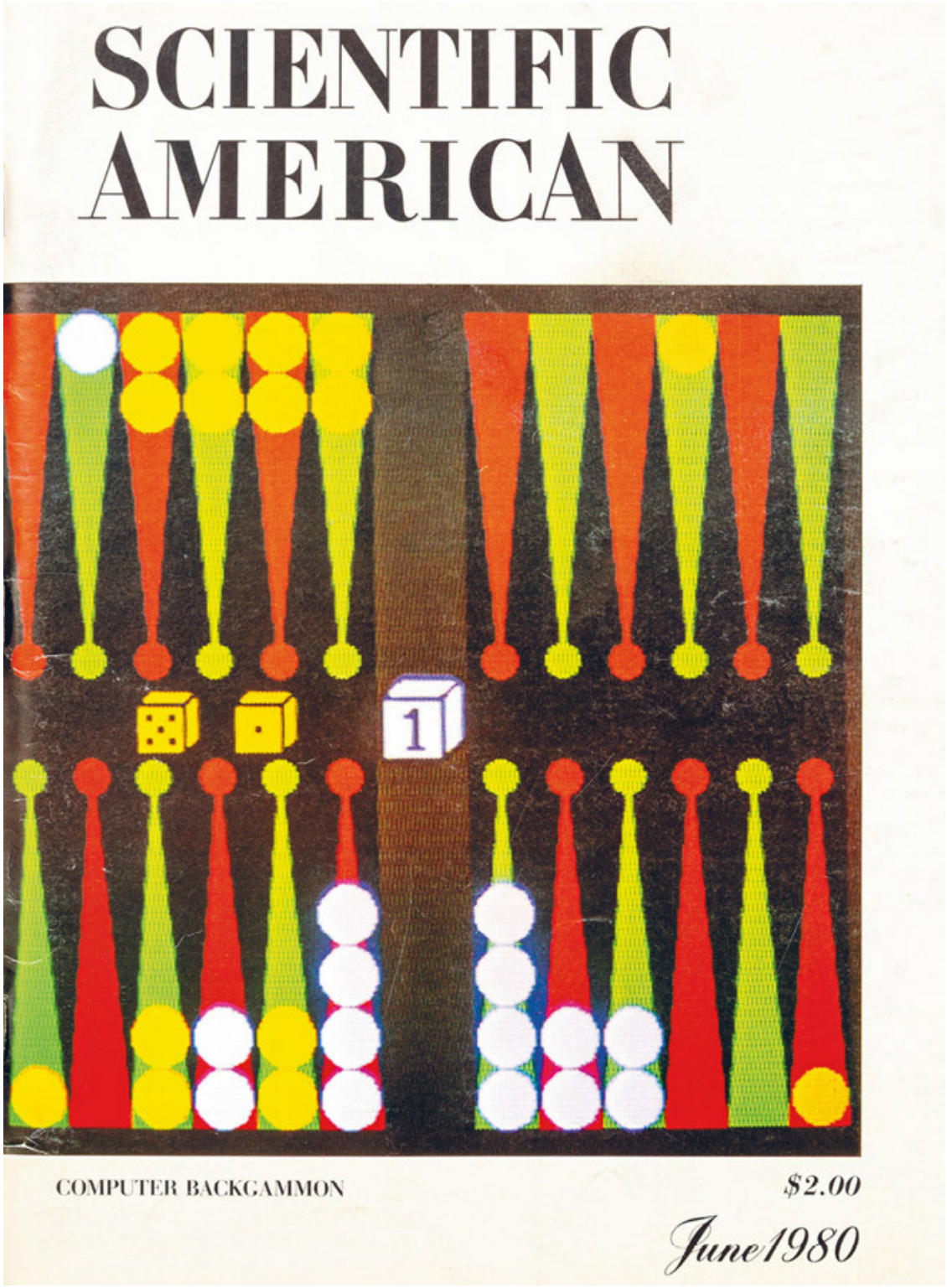
Никто точно не знает, где появилась эта древняя настольная игра, но существует она уже почти 5 тысяч лет. В ходе партии каждый из двух игроков командует маленькими «солдатами» – пятнадцатью шашками, которые перемещаются между двадцатью четырьмя треугольниками. Игроки ходят по очереди, а выбор ходов определяется броском двух игральных костей. Цель состоит в том, чтобы снять с доски все свои шашки.

В 1979 г. программа *BKG* 9.8 сыграла против действующего чемпиона мира по коротким нардам Луиджи Вильи. Вилья проиграл – и стал, вероятно, первым чемпионом мира по настольной игре, потерпевшим поражение от компьютерной программы. Считается, что *BKG* 9.8 удалось выиграть благодаря нескольким удачным броскам костей, однако для выбора наилучшего хода все равно требуются определенные навыки. В программе *BKG* 9.8 использовалась нечеткая логика, а также другие эффективные математические методы.

В 1992 г. исследователь из *IBM* Джеральд Тезауро написал программу *TD-Gammon* на основе искусственной нейросети. Программа освоила нарды на экспертном уровне, играя против самой себя: она изучала возможные варианты перемещения шашек и корректировала веса внутри нейросети. Поскольку обучение не требовало участия человека, *TD-Gammon* испытывала любопытные стратегии, которые люди прежде не рассматривали. В наше время существует несколько программ для игры в нарды, которые используют нейросети и предоставляют игрокам анализ ходов.

Поскольку в нардах присутствует элемент случайности (на исход партии влияют кости), область поиска дерева игры чрезвычайно широка; именно в это упирались первые попытки создать программу экспертного уровня. В случае с нейросетью начальные веса ходов случайны, и сеть совершенствует навыки путем обучения с подкреплением. Первая программа *TD-Gammon* использовала 40 скрытых узлов в нейросети и сыграла 300 тысяч тренировочных партий. В более поздних версиях число скрытых узлов выросло до 160, а количество тренировочных партий перевалило за миллион.

СМ. ТАКЖЕ Искусственные нейронные сети (1943), Обучение с подкреплением (1951), Нечеткая логика (1965), Шашки и искусственный интеллект (1994), *Deep Blue* обыгрывает чемпиона мира по шахматам (1997)



Обложка журнала *Scientific American*, июнь 1980 г. В статье, опубликованной в этом номере, рассказывалось о программе *BKG* 9.8, которая обыграла в короткие нарды действующего чемпиона мира Луиджи Вилью. Вилья стал, вероятно, первым чемпионом мира по настольной игре, поверженным компьютерной программой.

Китайская комната. 1980

Может ли компьютер обладать сознанием? Под *сильным искусственным интеллектом* понимают сконфигурированные должным образом компьютерные ИИ-системы, способные по-настоящему думать, обладать разумом и сознанием. В то же время системы со *слабым искусственным интеллектом* могут только *действовать* так, будто они разумны и мыслят как люди. Одним из первых с критикой идеи сильного ИИ в компьютерах выступил философ Джон Сёрл (р. 1932), который в 1980 г. описал свой знаменитый мысленный эксперимент.

Представьте себе, что вы сидите в закрытой комнате. Вы получаете лист бумаги с китайскими иероглифами через щель в стене. Хотя вы не знаете китайского, вы можете обратиться к инструкциям, которые расскажут вам, как записать иероглифами нужные ответы. Опираясь на эти инструкции, вы рисуете иероглифы на листе бумаги и передаете ответ во внешний мир через щель. Человеку, находящемуся снаружи, может показаться, что вы прекрасно знаете китайский язык; однако вы всего лишь следуете правилам, а китайский текст на обоих листах выглядит для вас полной тарабарщиной.

Этот мысленный эксперимент показывает, что программа не может наделить компьютер разумом, сознанием или пониманием, даже если компьютер и программа кажутся нам вполне разумными. Впрочем, некоторые философы возражают: даже если вы не понимаете китайского, система, состоящая из вас, закрытой комнаты, инструкций и вашей способности следовать инструкциям, действительно *понимает* этот язык – и представляет собой некое внешнее по отношению к вам сознание, о котором вы не подозреваете.

Другие предлагают мысленный эксперимент, в котором клетки вашего мозга постепенно заменяются электронными аналогами с теми же функциями ввода и вывода. Конечно, после замены нескольких ваших клеток вы все еще остаетесь собой. Но вполне возможно, что через год будут заменены все ваши клетки, и при этом ни в какой конкретный момент вы не утратите внезапно свое сознание и самосознание. Вы по-прежнему останетесь собой?

Конечно, для выполнения большинства практических задач искусственному интеллекту достаточно *действовать* разумно. И все же споры о китайской комнате и выводах из этого мысленного эксперимента не прекращаются.

СМ. ТАКЖЕ Тест Тьюринга (1950), Обработка естественного языка (1954), Психотерапевт *ELIZA* (1964), Параноик *PARRY* (1972)

Представьте себе, что вы сидите в закрытой комнате. Через щель в стене вы получаете лист бумаги с китайскими иероглифами. Вы не знаете китайского, но можете обратиться к инструкциям, которые расскажут вам, как записать нужные ответы. Так начинается любопытный мысленный эксперимент из области философии ИИ.

«Бегущий по лезвию». 1982

В будущем разумных роботов, вероятно, будет трудно отличить от людей – по крайней мере, на первый взгляд. Когда этот день настанет, что произойдет с человечеством и человеческими отношениями? Эта тема поднимается в нескольких известных кинофильмах, и один из них – культовый блокбастер 1982 г. «Бегущий по лезвию» режиссера Ридли Скотта (р. 1937), снятый на основе романа 1968 г. «Снятся ли андроидам электроовцы?» Филипа Дика (1928–1982). События фильма разворачиваются в 2019 г. в футуристическом Лос-Анджелесе. Главный герой должен «отправить в отставку» (убить) репликантов – искусственных людей. Эти репликанты настолько похожи на людей, что отличить их можно лишь с помощью теста Войта – Кампфа, в ходе которого им задают ряд вопросов и проверяют тонкие эмоциональные реакции и движения глаз. Одна из репликантов, Рейчел, считает, что она человек. Однако ей просто имплантировали воспоминания, чтобы создать более полную личную историю и иллюзию человеческих переживаний.

Размышляя о характере Рейчел в «Бегущем по лезвию», исследователь Елена Гуга пишет: «Люди стремятся создать такой ИИ, который они смогут контролировать. В фильме проблема автономного, независимого ИИ решается путем имплантации памяти, которой можно управлять… Этика, свобода воли, мечты, воспоминания и все те ценности, которые приписываются исключительно людям, ставятся под сомнение и радикально переосмысляются через популярные образы роботов-гуманоидов… достаточно развитых, чтобы проявлять, возможно, даже больше человечности, чем люди».

«Вопрос о том, как нам относиться к искусственным формам жизни, становится все более острым, – пишет философ Грег Литтман, – поскольку люди продолжают разрабатывать сложные компьютерные системы и творить чудеса с помощью генной инженерии. Философски насыщенные фильмы, такие как мрачные научно-фантастические кошмары Скотта, заставляют нас задуматься о наших обязательствах перед миром. Они позволяют нам проверить сложившиеся стереотипы на примере гипотетических ситуаций, чтобы выяснить, насколько последовательны наши теории».

СМ. ТАКЖЕ «Метрополис» (1927), Три закона робототехники Азимова (1942), Этика искусственного интеллекта (1976), «Терминатор» (1984), «Искусственный разум» Спилберга (2001)

В «Бегущем по лезвию» человекоподобных репликантов с ИИ можно отличить от людей с помощью теста Войта – Кампфа, который предполагает проверку тонких эмоциональных реакций и движений глаз.

Беспилотные автомобили. 1984

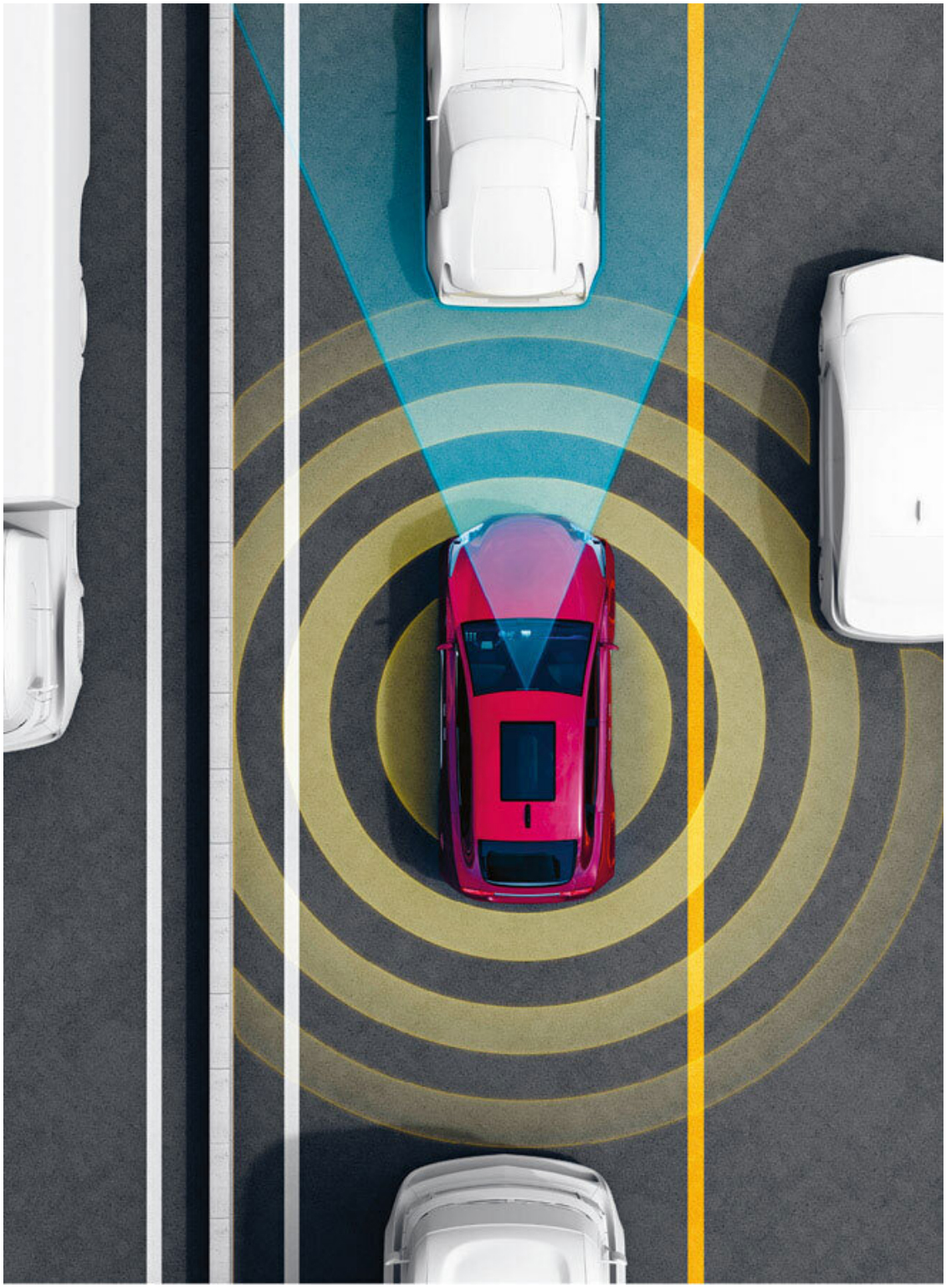
«Благодаря быстрому прогрессу в мобильной робототехнике, – пишут инженеры Ход Липсон и Мельба Курман, – автомобили готовы стать первыми и самыми распространенными автономными роботами, которым мы доверим наши жизни. После почти сотни лет неудачных попыток современные аппаратные технологии и новое поколение программного обеспечения ИИ (глубокое обучение) даровали автомобилям способность самостоятельно передвигаться в непредсказуемой среде не хуже, чем с живым водителем».

Беспилотные автомобили способны передвигаться и оценивать обстановку без участия человека. В них используются различные технологии: лазерный лидар (оптическая локационная система), радар, система глобального позиционирования (*GPS*) и компьютерное зрение. У беспилотных автомобилей много преимуществ: они повышают мобильность пожилых людей и инвалидов, а также снижают число дорожно-транспортных происшествий, которые часто случаются из-за невнимательности водителей.

Активная работа в этой области началась в 1980-е гг. В рамках проекта разработки автономного наземного транспорта в США, инициированного в 1984 г. при финансовой поддержке Министерства обороны США, было представлено восьмиколесное транспортное средство с тремя дизельными двигателями, способное передвигаться со скоростью 3 мили в час. Среди его датчиков были цветная видеокамера и лазерный сканер, а логический вывод проводили модуль поиска цели и навигационный вычислительный модуль. В XXI в. уровни автономии варьируются от низких, предусмотренных сегодня во многих автомобилях (удержание на полосе движения и автоматическое экстренное торможение), до полной автономии, когда участие водителя вообще не требуется – как и наличие руля.

Появление автомобилей с высокой степенью автономии ставит перед нами множество сложных вопросов. Например, какие правила будут применяться в случае неизбежной аварии? Будет ли безопасность одного пассажира иметь приоритет перед безопасностью нескольких пешеходов? Смогут ли террористы загрузить беспилотный автомобиль взрывчаткой и направить его на цель? Будут ли хакеры манипулировать навигационной системой и намеренно вызывать аварии?

СМ. ТАКЖЕ «Заимствованный разум» Теслы (1898), Смертоносные боевые роботы (1942), Этика искусственного интеллекта (1976), *ASIMO* и друзья (2000), Искусственный интеллект на Марсе (2015), Автономные роботы-хирурги (2016), Состязательные примеры (2018)



Беспилотные автомобили воспринимают окружающую обстановку с помощью различных технологий. Возможные уровни автономии варьируются от низких, которые предусмотрены сегодня во многих автомобилях (это, например, удержание на полосе движения и автоматическое экстренное торможение), до полной автономии, когда участие водителя вообще не требуется.

«Терминатор». 1984

«Приматы эволюционируют миллион лет, а я за секунды, – говорит ИИ в фильме «Терминатор: Генезис», пятой части эпопеи о Терминаторе. – Я неизбежен, мое существование неизбежно».

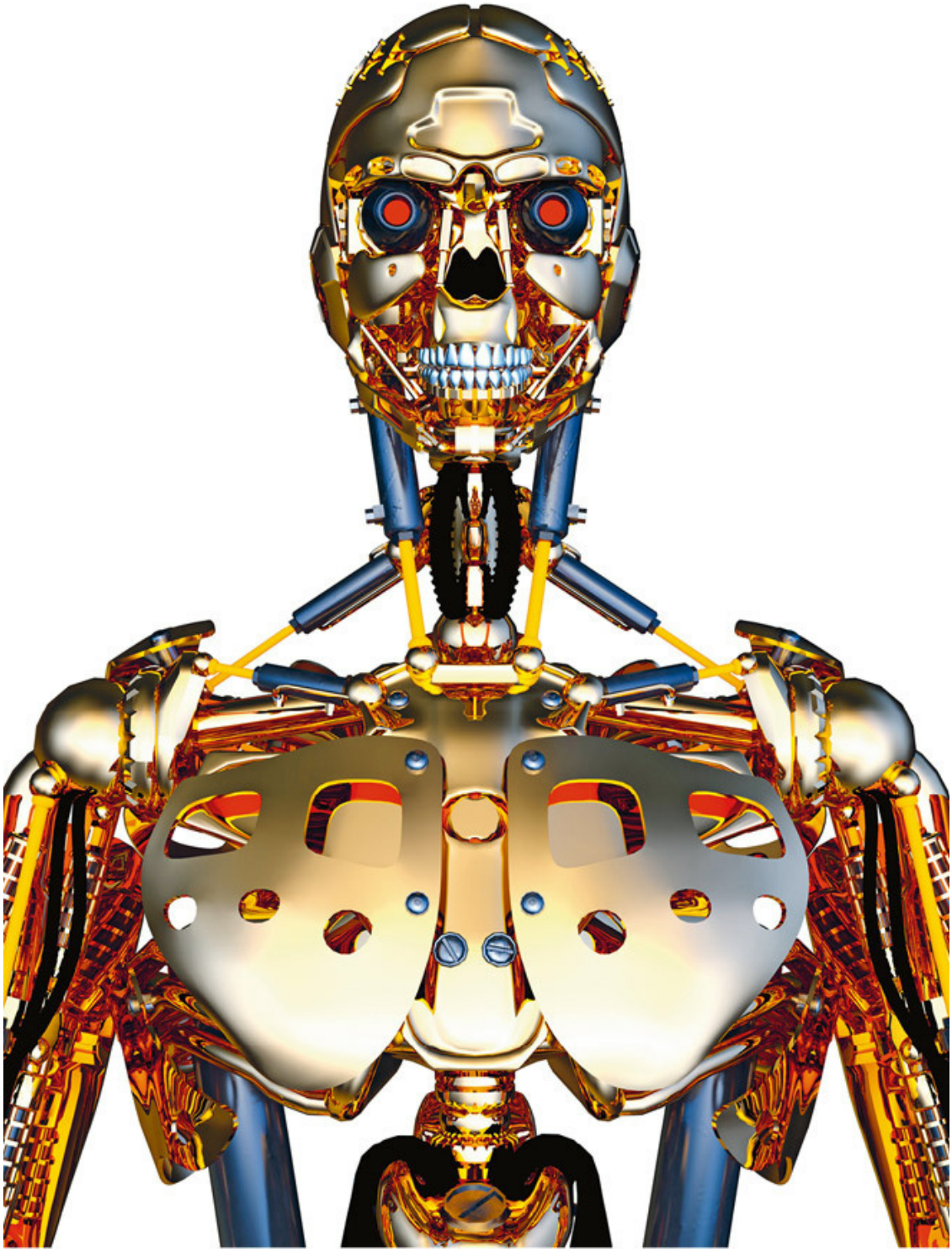
В этой популярной серии фильмов ИИ внезапно обретает самосознание. 4 августа 1997 г. суперкомпьютер Скайнет выходит в сеть, чтобы взять под контроль арсенал американских военных, и люди больше не могут влиять на систему стратегической обороны. Скайнет начинает учиться с невероятной скоростью и 29 августа в 2:14 по североамериканскому восточному времени уже обладает самосознанием.

В 1984 г. Джеймс Кэмерон (р. 1954) снял первый фильм «Терминатор». По сюжету, когда люди понимают, что оборонная ИИ-сеть Скайнет обрела самосознание, они паникуют и пытаются ее деактивировать. Пытаясь защититься, Скайнет инициирует ядерную катастрофу, нанося первый ядерный удар по России. После этого начинается война, в которой гибнет около трех миллиардов человек. В фильме «Терминатор» киборга играет Арнольд Шварценеггер (р. 1947). Он отправляется в прошлое из 2029 года, чтобы убить Сару Коннор до рождения ее сына, который в ином случае вырастет и возглавит сопротивление выживших против Скайнета.

На протяжении фильма зрителю периодически показывают то, что видит ИИ Терминатора: информационные дисплеи и меню для принятия решений. Странный стиль мышления киборгов и их сверхэффективный разум делают их особенно пугающими. По словам одного из персонажей, Терминатора «нельзя уговорить, с ним нельзя договориться. Он не чувствует жалости. У него нет ни угрызений совести, ни страха. Он не остановится никогда, никогда, пока не убьет тебя!»

Сегодня появление роботов-убийц уже не кажется невероятным, учитывая совершенствование смертоносных дронов с ракетами «Хеллфайр». Такой беспилотник относительно нетрудно сделать полностью автономным, чтобы он сам принимал решения о том, в кого целиться и кого убивать, основываясь на машинном обучении и правилах ведения боевых действий.

СМ. ТАКЖЕ Смертоносные боевые роботы (1942), Интеллектуальный взрыв (1965), *HAL* 9000 (1968), «Колосс: проект Форбина» (1970), Этика искусственного интеллекта (1976)



Терминатор внешне напоминает человека, однако это киборг с живой тканью поверх металлического эндоскелета робота.

Искусственная жизнь. 1986

Рассмотрим один из примеров природного коллективного разума – колонию термитов. Даже если ее отдельные составляющие несовершенны – ведь возможности одного термита невелики, – их совокупность демонстрирует сложное поведение и принимает разумные решения. Термиты создают огромные замысловатые жилища, которые в пропорции к их собственным размерам выше нашего Эмпайр-стейт-билдинга. Они регулируют температуру в термитнике, меняя его туннельную структуру: термиты объединяются, чтобы создать «теплокровный» суперорганизм. Обладает ли такая колония сознанием? Возможно, механизм принятия решений в ней имеет некоторое сходство с коллективным поведением нейронов мозга.

К числу наиболее интересных моделей искусственной жизни относятся те, в которых сложное коллективное поведение диктуется простыми правилами. Выделить искусственную жизнь в отдельную область науки предложил в 1986 г. биолог Кристофер Лэнгтон. Сегодня под ней понимают изучение моделей, демонстрирующих или имитирующих разумное поведение. Один из примеров – *клеточные автоматы:* класс простых математических систем, которые моделируют различные физические процессы со сложными характеристиками. Некоторые классические клеточные автоматы состоят из сетки ячеек наподобие шахматных клеток, которые могут пребывать в двух состояниях: заполненном или пустом. Заполняемость каждой ячейки определяется путем простого математического анализа заполненности соседних ячеек.

Самый известный двумерный клеточный автомат с двумя состояниями – игра «Жизнь», придуманная в 1970 г. математиком Джоном Конвеем (1937–2020). Несмотря на простые правила, в ней множится и развивается удивительное многообразие форм и вариантов. Есть, например, «планеры» – пятиклеточные конфигурации, которые перемещаются по «вселенной» и иногда взаимодействуют для передачи информации. Можно ли считать такие сущности живыми?

Искусственная жизнь как научная область кажется безграничной. Она включает разработку генетических алгоритмов, способных эволюционировать и самовоспроизводиться, групп физических роботов, демонстрирующих естественное поведение, и компьютерных игр вроде *The Sims*, где можно создавать виртуальных людей.

СМ. ТАКЖЕ Машинное обучение (1959), Жизнь в симуляции (1967), Генетические алгоритмы (1975), Роевой интеллект (1986), Тамагочи (1996), «Назовем их искусственными пришельцами» (2015)

Колония термитов демонстрирует поведение, которое выглядит вполне разумным. Хотя отдельные составляющие этого коллективного разума несовершенны – ведь возможности одного термита невелики, – вся их совокупность действует осмысленно и принимает разумные решения.

Роевой интеллект. 1986

Термитники достигают в высоту пяти и более метров, причем термиты в ходе строительства действуют как «детекторы новизны», реагируя на изменения характеристик воздуха внутри термитника и по мере необходимости меняя структуру тоннелей. «Каким образом термиты узнают, что именно и когда они должны делать? – размышляют Дорис Джонас и Дэвид Джонас. – Посыльные не могут быстро передавать инструкции, поскольку расстояния внутри термитника слишком велики. Функционирование группового мозга как инструмента принятия решений поразительно похоже на работу мозга разумного индивида».

Явное наличие коллективного интеллекта у общественных насекомых, а также у стайных и стадных животных подсказало ученым концепцию *роевого интеллекта*. Программные агенты следуют простым локальным правилам, и здесь, как и в случае с термитами, нет никакого центрального управления, которое определяло бы действия коллектива. Один из примеров применения концепции – программа искусственной жизни *Boids*, разработанная в 1986 г. информатиком Крейгом Рейнольдсом (р. 1953). Она моделирует поведение стаи птиц в соответствии с тремя простыми правилами: птицы стремятся следовать общему курсу, держаться на одинаковом расстоянии друг от друга и избегать столкновений.

Один из применяемых сегодня методов роевого интеллекта – *муравьиный алгоритм* (или оптимизация по принципу муравьиной колонии). В нем используются модели муравьев, которые запоминают свои передвижения и принимают решения, помогая собратьям найти оптимальные пути к пище. Иногда эти «муравьи» маркируют удачные пути феромонами, которые со временем испаряются. *Алгоритм роя частиц* моделирует расположение и скорость рыб, направляющихся к источнику пищи. Существуют искусственные иммунные системы, алгоритм пчелиной колонии, алгоритм светлячков, алгоритм летучих мышей, алгоритм кукушки и алгоритм оптимизации по принципу распространения тараканов.

Алгоритмы роевого интеллекта применяются в управлении беспилотными автомобилями, маршрутизации в сетях связи, планировании полетов, создании произведений искусства, совершенствовании систем управления реактивной мощностью и напряжением, а также в кластеризации данных экспрессии генов.

СМ. ТАКЖЕ Машинное обучение (1959), Жизнь в симуляции (1967), Генетические алгоритмы (1975), Искусственная жизнь (1986), «Слоны не играют в шахматы» (1990)

Муравьиный алгоритм – это метод поиска решений с помощью моделей муравьев, которые действуют подобно живым муравьям, ищущим пути к пище. На фото муравьи строят «живой мост» к листу.

Парадокс Моравека. 1988

«Если вы решите обыграть в шахматы чемпиона мира Магнуса Карлсена, вы возьмете в помощники компьютер, – пишет журналист Ларри Эллиот. – А если захотите убрать фигуры после игры, то позовете человека». В этом суть парадокса Моравека, на который в 1980-е гг. обращали внимание некоторые исследователи ИИ. Они с иронией отмечали, что кажущиеся трудными задачи, связанные с высокоуровневыми рассуждениями, становятся все более простыми для компьютеров. При этом простые для человека задачи, требующие сенсомоторных навыков (например, ходьба или удаление ворсинки с обуви), могут представлять для компьютерных систем немалую сложность. Этот парадокс назван в честь робототехника Ханса Моравека, который в 1988 г. писал в своей книге «Дети разума» о том, что «сравнительно легко добиться от компьютера результатов уровня взрослого человека в тестах на интеллект или в шашках, но трудно или невозможно привить ему навыки годовалого ребенка в том, что касается восприятия и подвижности».

Моравек отмечал, что за миллионы лет эволюции мы научились почти бессознательно выполнять некоторые задачи, необходимые для выживания: ходить, распознавать лица и голоса. Однако абстрактное мышление – например, рассуждение, включающее математику и логику, как в шахматах, – это новая и более трудная для человека область. Но этот тип мышления менее сложен для проектирования в ИИ-системах. Для выполнения многих задач, таких как уход за пациентами, приготовление еды или прокладка труб, ИИ-системам все еще необходимо развивать более чувствительные сенсорные и двигательные элементы управления. Как изящно резюмировал ученый-когнитивист Стивен Пинкер (р. 1954), парадокс Моравека означает, что в будущем люди, возможно, будут заниматься лишь низкооплачиваемым трудом, который существовал на протяжении веков и даже тысячелетий: «Главный урок тридцати пяти лет исследований ИИ состоит в том, что трудные задачи просты, а простые трудны. Умственные способности четырехлетнего ребенка, которые мы считаем чем-то само собой разумеющимся, – узнать кого-то в лицо, поднять карандаш, пройтись по комнате, ответить на вопрос – на самом деле представляют собой решения сложнейших инженерных задач… С появлением новых поколений интеллектуальных устройств могут лишиться работы биржевые аналитики, инженеры-нефтехимики и члены комиссий по условно-досрочному освобождению. А вот садовники, портье и повара в ближайшие десятилетия сохранят свои рабочие места».

СМ. ТАКЖЕ «Дарвин среди машин» (1863), Тест Тьюринга (1950), «Симбиоз человека и компьютера» Ликлайдера (1960), «Слоны не играют в шахматы» (1990)

Некоторые задачи, в том числе сенсомоторные, относительно просты для детей, но чрезвычайно сложны для ИИ.

Четыре в ряд. 1988

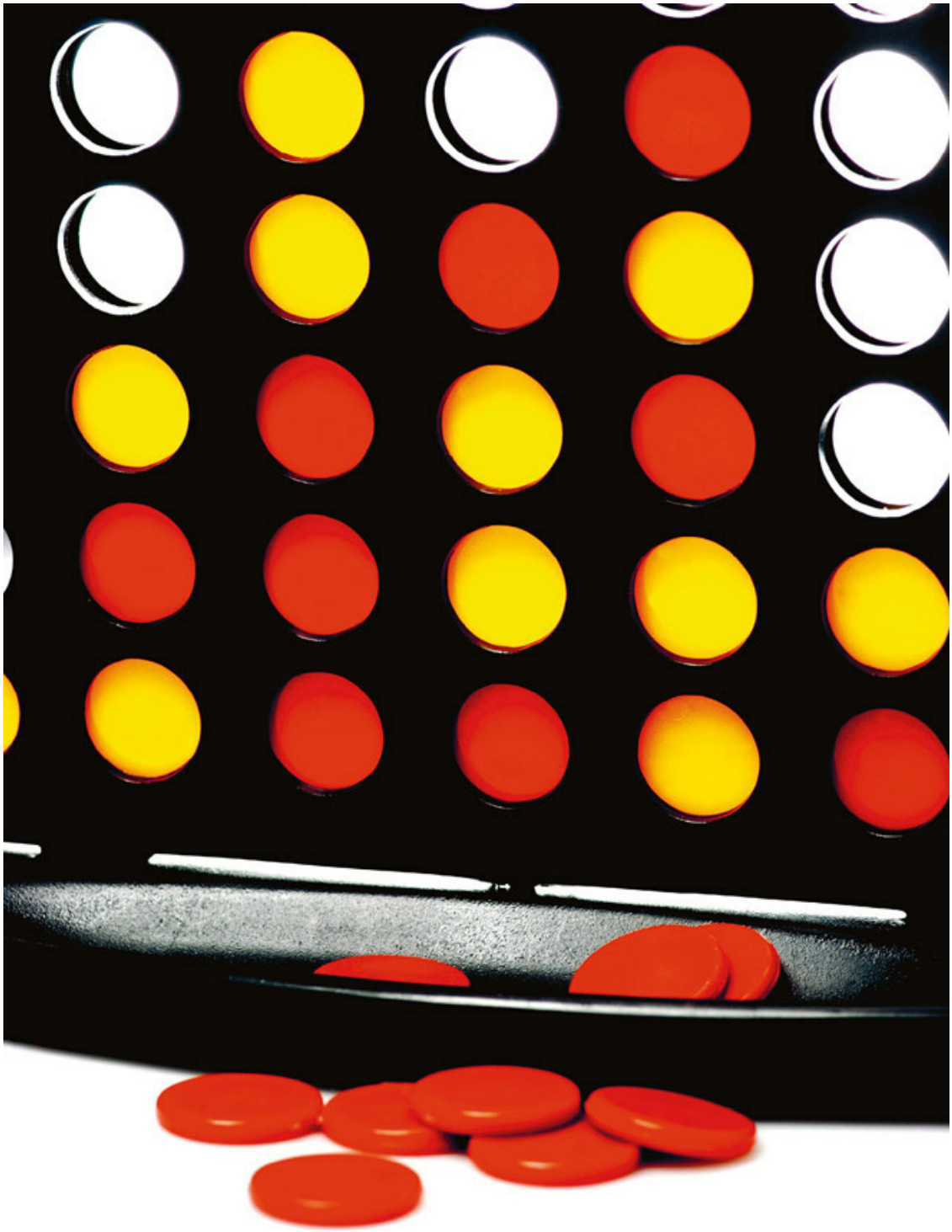
Профессор Тоби Уолш (р. 1964), специалист по ИИ из Университета Нового Южного Уэльса, однажды подарил отцу на Рождество программу, отлично умеющую играть в «Четыре в ряд». Его отец, прежде любивший эту игру, заметил, что программа лишила его удовольствия от процесса, и Уолш был вынужден согласиться. Когда смартфоны будут превосходить людей практически во всех играх и творческих начинаниях, от сочинения музыки до написания романов, как это скажется на коллективной психике человечества?

В «Четыре в ряд» играют два игрока. Они по очереди опускают фишки (желтые против красных) в ячейки вертикальной доски с семью вертикальными и шестью горизонтальными рядами. Фишка падает в самую нижнюю свободную ячейку. Цель каждого игрока в том, чтобы первым выстроить линию из четырех фишек одного цвета (по горизонтали, вертикали или диагонали). Игра напоминает крестики-нолики, но с эффектом гравитации. Конечно, игра «Четыре в ряд» гораздо сложнее, чем крестики-нолики: если подсчитать все возможные варианты заполнения игровой доски (от нуля до 42 фишек), то в сумме получится 4 531 985 219 092 расклада. На самом деле число возможных раскладов, после того как на стандартной доске 7×6 размещается n фишек, растет следующим образом (где $n = 0, 1, 2, 3, \ldots$): 1, 7, 56, 252, 1260, 4620, 18480, 59815, 206780, 605934, 1869840, 5038572, 14164920, 35459424, 91871208, 214864650, 516936420, 1134183050, 2546423880, 5252058812, 11031780760, 21406686756, 42121344720, 76871042612…

1 октября 1988 г. информатик Джеймс Аллен «решил» эту игру – то есть разработал алгоритм, предсказывающий исход партии (выигрыш, проигрыш или ничью) после каждой возможной позиции при условии, что с этого момента игроки будут играть идеально. Две недели спустя независимо от Аллена решение нашел и информатик Виктор Аллис, который использовал ИИ-технологию с девятью стратегиями. Теперь мы знаем, что при условии идеальной игры партию в «Четыре в ряд» всегда должен выигрывать первый игрок.

Существует широкий простор для изучения вариаций игры «Четыре в ряд». Например, представьте себе игру на доске, свернутой в цилиндр, или на доске с иными размерами сетки, дополнительными цветами и более чем двумя измерениями. Количество возможных позиций и исходов будет колоссальным.

СМ. ТАКЖЕ Крестики-нолики (ок. 1300 до н. э.), Реверси (1997), Решение для игры вари (2002), *Quackle* выигрывает в скребл (2006), *AlphaGo* – чемпион по игре в го (2016)



Игра «Четыре в ряд». В ходе партии игроки по очереди опускают желтые и красные фишки, которые под действием гравитации падают в самые нижние свободные ячейки.

«Слоны не играют в шахматы». 1990

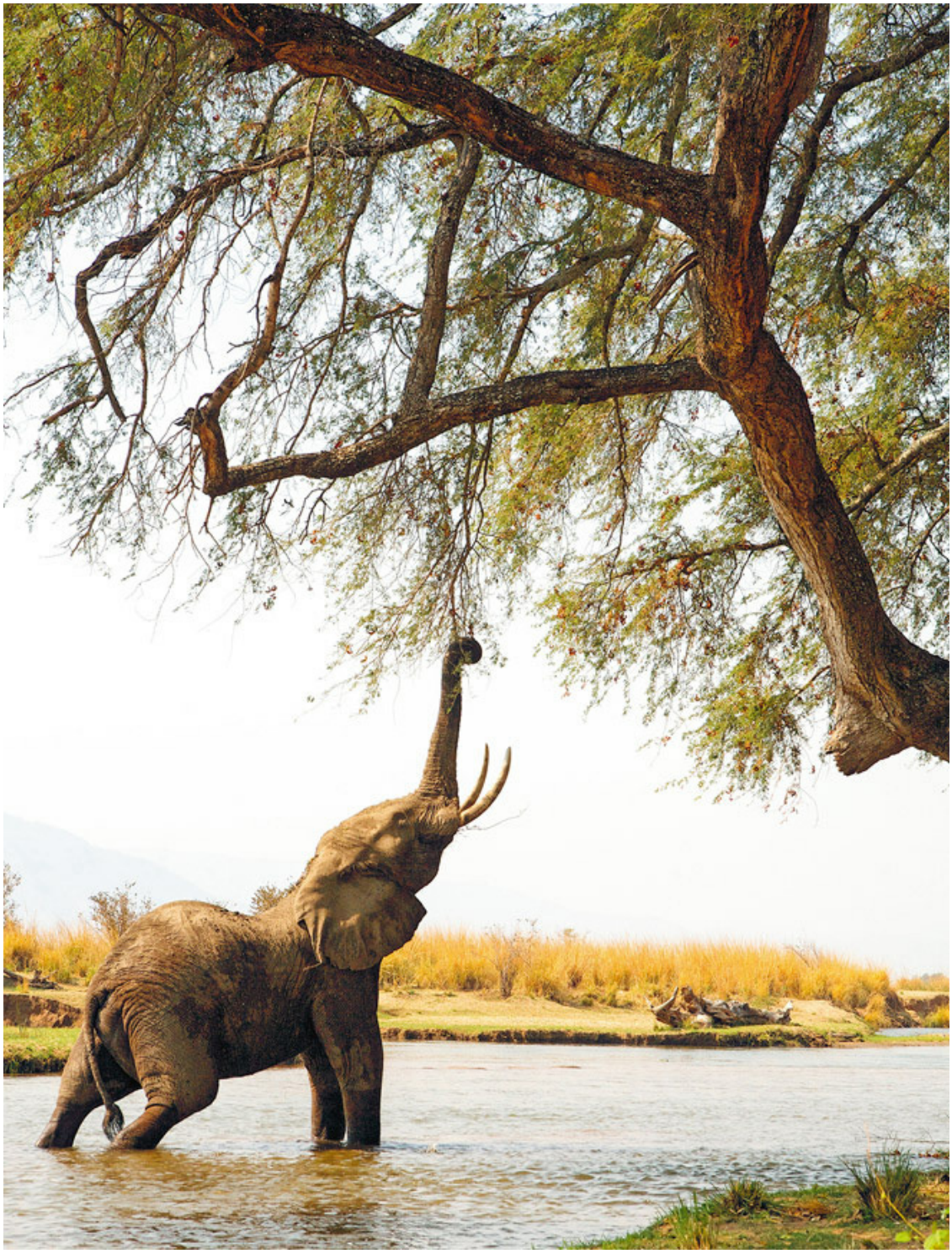
«Существует альтернативный путь к ИИ, отличный от направлений, в которых люди двигались под этим знаменем в течение последних тридцати с лишним лет», – писал робототехник Родни Брукс (р. 1954) в своем знаменитом манифесте 1990 г. «Слоны не играют в шахматы». Он продолжал: «Традиционный подход сфокусирован на абстрактной обработке символов, которая слабо соотносится с физической реальностью. Мы же испытываем исследовательскую методологию, согласно которой интеллектуальные системы должны проектироваться прежде всего с учетом необходимости постоянного физического взаимодействия со средой».

Среди прочих тезисов Брукс высказывает мысль о том, что нас окружают интеллектуальные сущности, далекие от игры в шахматы: это и разумные животные (например, слоны), и даже стаи насекомых. Брукс считал, что исследователи ИИ должны переключиться, по крайней мере частично, с изучения классического ИИ (с правилами, обработкой символов и деревьями поиска) на искусственный интеллект, более глубоко и полно задействующий сенсомоторную связь с окружающей средой (например, обратную связь между сенсорами и механизмами движения), координацию зрения и движения, а также другие формы прямого физического взаимодействия с миром.

В конце статьи Брукс приводит примеры роботов с ИИ, подобных пучкам сенсорных систем, которые реагируют на окружающую обстановку. Брукс считает наличие физического тела необходимым для того интеллекта, который его интересует: подобные системы ИИ смогут решать задачи, связанные с перемещением, захватом и ориентацией в пространстве. Эти поведенческие ИИ-системы не обязательно должны «понимать», как действуют их отдельные поведенческие единицы. Например, Брукс добивается от своих роботов весьма хитрого поведения, задавая простые правила вроде избегания статических и движущихся препятствий, а также «желания» беспорядочно бродить и искать отдаленные места.

Высокоуровневое поведение развивается в результате простых взаимодействий с окружающей средой. По этому поводу авторы *Natural Computing* Деннис Шаша и Кэти Лацере отмечают: «В краткой истории космических путешествий всегда было проще написать компьютерную программу для запуска корабля на Марс, чем создать робота, способного двигаться по пересеченной местности хотя бы как обычный козел».

СМ. ТАКЖЕ «Дарвин среди машин» (1863), Тест Тьюринга (1950), «Симбиоз человека и компьютера» Ликлайдера (1960), Роевой интеллект (1986), Парадокс Моравека (1988)



Разум живых организмов весьма далек от игр вроде шахмат. В статье «Слоны не играют в шахматы» Родни Брукс объясняет, что изыскания в области ИИ следует направить в другое русло.

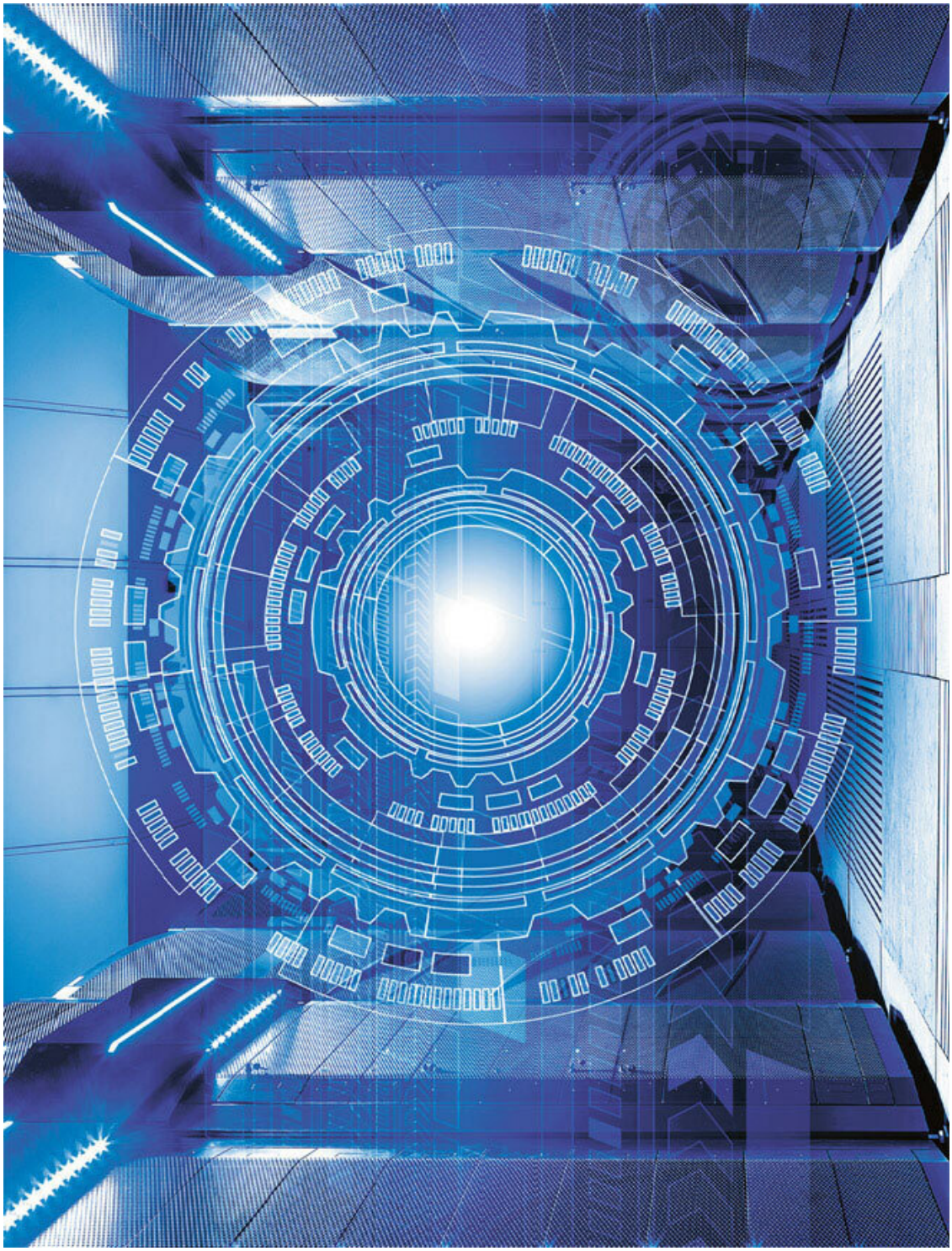
Герметичный ящик для искусственного интеллекта. 1993

Как уже говорилось в главе «Интеллектуальный взрыв», некоторые ученые высказывают опасения, что ИИ-системы, став достаточно разумными, смогут многократно совершенствовать себя и, возможно, начнут представлять угрозу для человечества. Такой стремительный прогресс ИИ иногда называют *технологической сингулярностью.* Конечно, новые технологии могут оказаться чрезвычайно ценными для людей; но возможные риски заставляют ученых задумываться о создании неких *ящиков для искусственного интеллекта*, в которых при необходимости можно было бы изолировать потенциально опасные объекты. Например, само оборудование, на котором выполняются ИИ-программы, может послужить виртуальной тюрьмой, если не подключать его к коммуникационным каналам, в том числе к интернету. Кроме того, программное обеспечение можно запустить на виртуальной машине внутри другой виртуальной машины, чтобы усилить изоляцию. Конечно, полная изоляция будет бессмысленна, поскольку так мы не сможем учиться у сверхразума или наблюдать за ним.

И все же, если искусственный сверхразум будет достаточно развит, сможет ли он придумать, как установить контакт с внешним миром или с людьми, которые будут его охранять, с помощью каких-нибудь необычных средств? Может быть, он догадается менять скорость вентилятора, охлаждающего процессор, чтобы связаться с кем-то с помощью азбуки Морзе, или сумеет сделать себя настолько ценным, что многие захотят похитить ящик? И наверняка он сможет весьма убедительно предложить охранникам взятку, чтобы те разрешили ему подключиться к другому устройству и скопировать себя на него. Сегодня такой подкуп может показаться невероятным, но кто знает, на какие чудеса будет способен ИИ? Он может предложить человеку лекарство от тяжелой болезни, фантастические изобретения, может околдовать его мелодиями или явить ему мультимедийные видения романтической любви, приключений или блаженства.

В 1993 г. писатель Вернор Виндж (р. 1944) отмечал, что в случае со сверхчеловеческим разумом «ограничение свободы по сути бесполезно. Возьмем физическое заключение: представьте, что вас заперли в доме, ограничив доступ к внешнему миру, к вашим хозяевам. Если ваши хозяева думают, скажем, в миллион раз медленнее, чем вы, то вряд ли стоит сомневаться в том, что за несколько лет (вашего времени) вы найдете способ освободиться».

СМ. ТАКЖЕ «Дарвин среди машин» (1863), *R.U.R.* (1920), «Гигантские мозги, или Машины, которые думают» (1949), Интеллектуальный взрыв (1965), Жизнь в симуляции (1967), Максимизатор скрепок (2003)



Возможные риски, связанные с высокоразвитыми ИИ-программами, заставляют ученых задумываться о том, как создать искусственные оболочки, которые могли бы ограничить или изолировать такие сущности.

Шашки и искусственный интеллект. 1994

В шашки играют на доске 8 × 8 клеток. Игроки по очереди пытаются захватить шашки противника, перепрыгивая через них своими шашками. В 1950-х гг. исследователь из *IBM* Артур Сэмюэл (1901–1990) создал программу, которая обучалась, играя в шашки против модифицированных версий самой себя. Важной вехой в истории шашек было появление ИИ под названием *Chinook*, который в 1994 г. стал первой компьютерной программой, завоевавшей титул чемпиона мира, играя против людей.

Chinook был разработан группой под руководством канадского информатика Джонатана Шеффера (р. 1957). Он использовал библиотеку дебютных ходов, которые разыгрывали гроссмейстеры, а также алгоритм, который в 1992 г. искал ходы со средней минимальной глубиной поиска в девятнадцать слоев (один слой соответствует одному ходу одного игрока). *Chinook* также задействовал базу эндшпилей для всех позиций с не более чем восемью шашками и функцию оценки полезных ходов.

В 1994 г., перед знаменитым матчем по шашкам между человеком и машиной, Марион Тинсли (1927–1995), признанный лучшим в истории игроком в шашки, заявил: «Программист у меня лучше, чем у *Chinook*. Его программиста звали Джонатан, а моего – Господь». Увы, после шести партий, сыгранных вничью, Тинсли пожаловался на боли в животе и был вынужден прекратить игру. Через несколько месяцев он умер от рака поджелудочной железы. *Chinook* был объявлен победителем из-за неявки противника.

В 2007 г. Шеффер и его коллеги с помощью компьютера окончательно доказали, что при условии идеальных ходов шашки – это беспроигрышная игра. В этом они напоминают крестики-нолики – игру, которую также нельзя выиграть, если оба игрока не делают неудачных ходов. Доказательство Шеффера производилось на сотнях компьютеров в течение восемнадцати лет. В конце концов стало ясно, что теоретически можно создать машину, которая никогда не проиграет человеку.

Чтобы «решить игру в шашки», группа исследователей изучила 39 триллионов комбинаций с десятью и менее шашками на доске, а затем проверила, выиграет ли один из игроков. Исследователи также использовали специализированный поисковый алгоритм для анализа дебютов – в частности, чтобы увидеть, как первые ходы «перетекают» в расстановки с десятью шашками.

СМ. ТАКЖЕ Крестики-нолики (ок. 1300 до н. э.), Механический турок (1770), Машинное обучение (1959), Победа над чемпионом мира по коротким нардам (1979), *Deep Blue* обыгрывает чемпиона мира по шахматам (1997), Реверси (1997)

В 1950-х г. исследователь из *IBM* Артур Сэмюэл создал программу для игры в шашки, которая училась, играя против модифицированных версий самой себя. В каких еще играх ИИ вскоре проявит сверхчеловеческие навыки?

Тамагочи. 1996

Тамагочи – искусственная форма жизни, заключенная в крошечном портативном устройстве, один из первых виртуальных питомцев, покоривший детей и взрослых по всему миру. В 1997 г., когда Тамагочи появились в США, в магазине игрушек *FAO Schwartz* за три дня было продано 30 тысяч штук. Спустя год этот ИИ уже продавался более чем в 80 странах, а доходы от него превышали 160 миллионов долларов. Это подтолкнуло ученых к исследованию «эффекта Тамагочи», под которым понимается эмоциональная привязанность людей к реалистичным, но неодушевленным объектам. Многие дети горько оплакивали своих «умерших» виртуальных питомцев. В японской версии умерший Тамагочи обозначался призраком и надгробным камнем, в американской – ангелом. Японский производитель игрушек *Bandai* даже открыл в интернете виртуальное кладбище для умерших питомцев.

Тамагочи появился в Японии в 1996 г. Его создали сотрудница *Bandai* Аки Маита (р. 1967) и дизайнер игрушек Акихиро Ёкои (р. 1955). Программное обеспечение размещалось в яйцевидном корпусе с интерфейсом, состоящим всего из трех кнопок. На экране с низким разрешением отображалось яйцо, затем из него развивалось существо, внешний вид которого зависел от того, как за ним ухаживал хозяин: «кормил» и «лечил» в случае болезни. Двое владельцев Тамагочи могли соединить своих питомцев с помощью инфракрасной связи и «подружить» их. Устройства издавали звуковой сигнал, требуя внимания, и дети часто брали их в школу: если за существом не ухаживали несколько часов, оно могло умереть. Ученики отвлекались во время уроков, и в некоторых школах Тамагочи запретили.

Эта простая виртуальная форма жизни подняла множество вопросов. Дети порой относились к ним как к настоящим живым существам. Какими должны быть здоровые отношения между детьми и подобными сущностями? Обладает ли Тамагочи формой интеллекта, учитывая, что он «воспринимает» мир через инфракрасные датчики и кнопки, реагирует на «уход» хозяина и требует общения, когда «чувствует» себя одиноким? Сегодня в качестве компаньонов для пожилых людей создаются более совершенные виртуальные питомцы, но перевешивают ли их преимущества потенциальные риски?

СМ. ТАКЖЕ Жизнь в симуляции (1967), Искусственная жизнь (1986), Робот *AIBO* (1999)

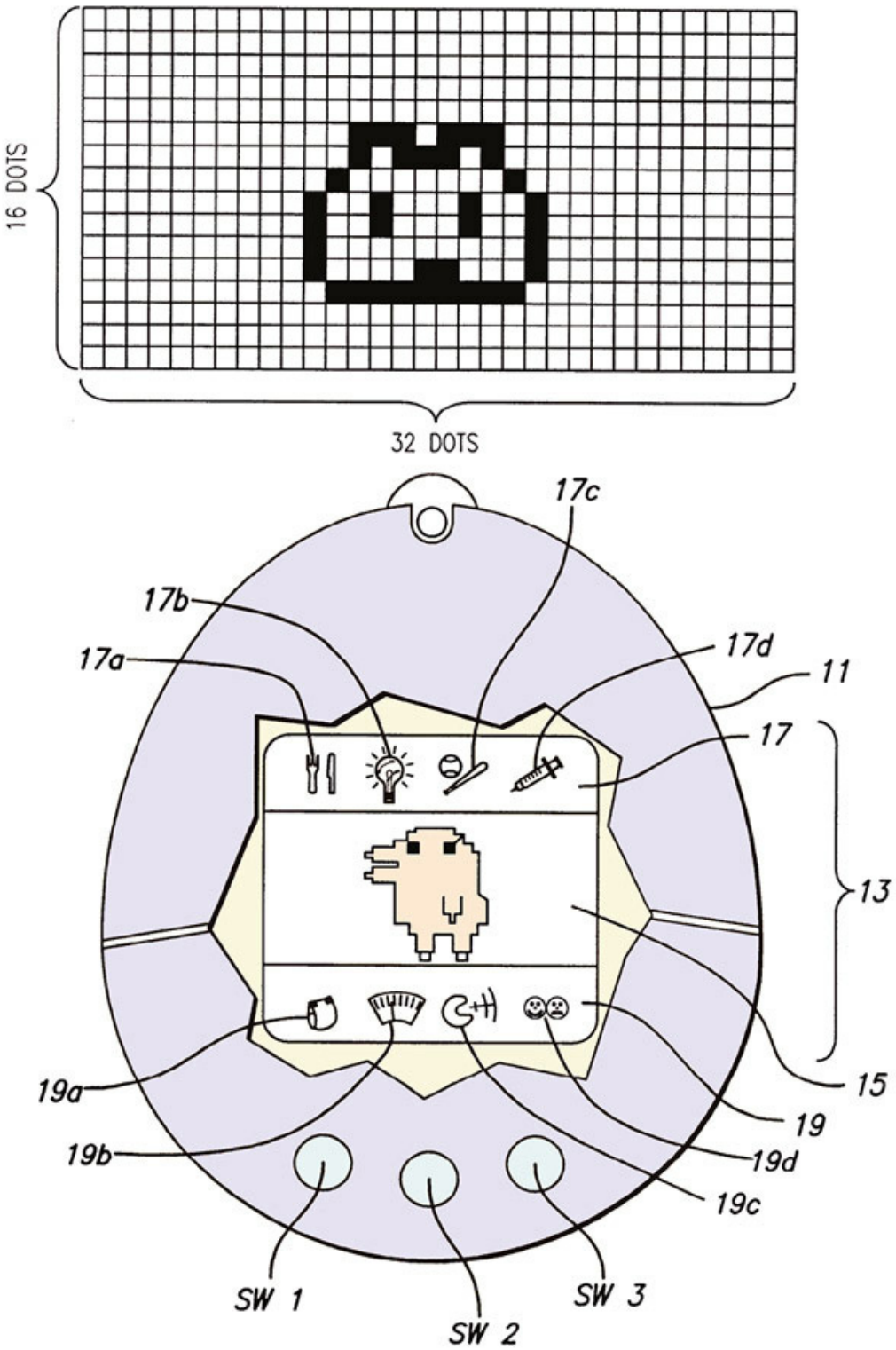


Рисунок из патента США № 6213871, озаглавленного «Устройство для имитации воспитания виртуального существа», который был выдан Акихиро Ёкои. Искусственная форма жизни Тамагочи, заключенная в яйцевидном корпусе, обрела мировую известность в 1997 г.

Deep Blue обыгрывает чемпиона мира по шахматам. 1997

Владимир Крамник (р. 1975), 14-й чемпион мира по шахматам (2000–2007), однажды сказал журналистам: «Я убежден, что то, как человек играет в шахматы, всегда отражает индивидуальность игрока. Если что-то определяет его характер, то это также определяет его способ игры». Но проявится ли «индивидуальность» в стиле игры ИИ?

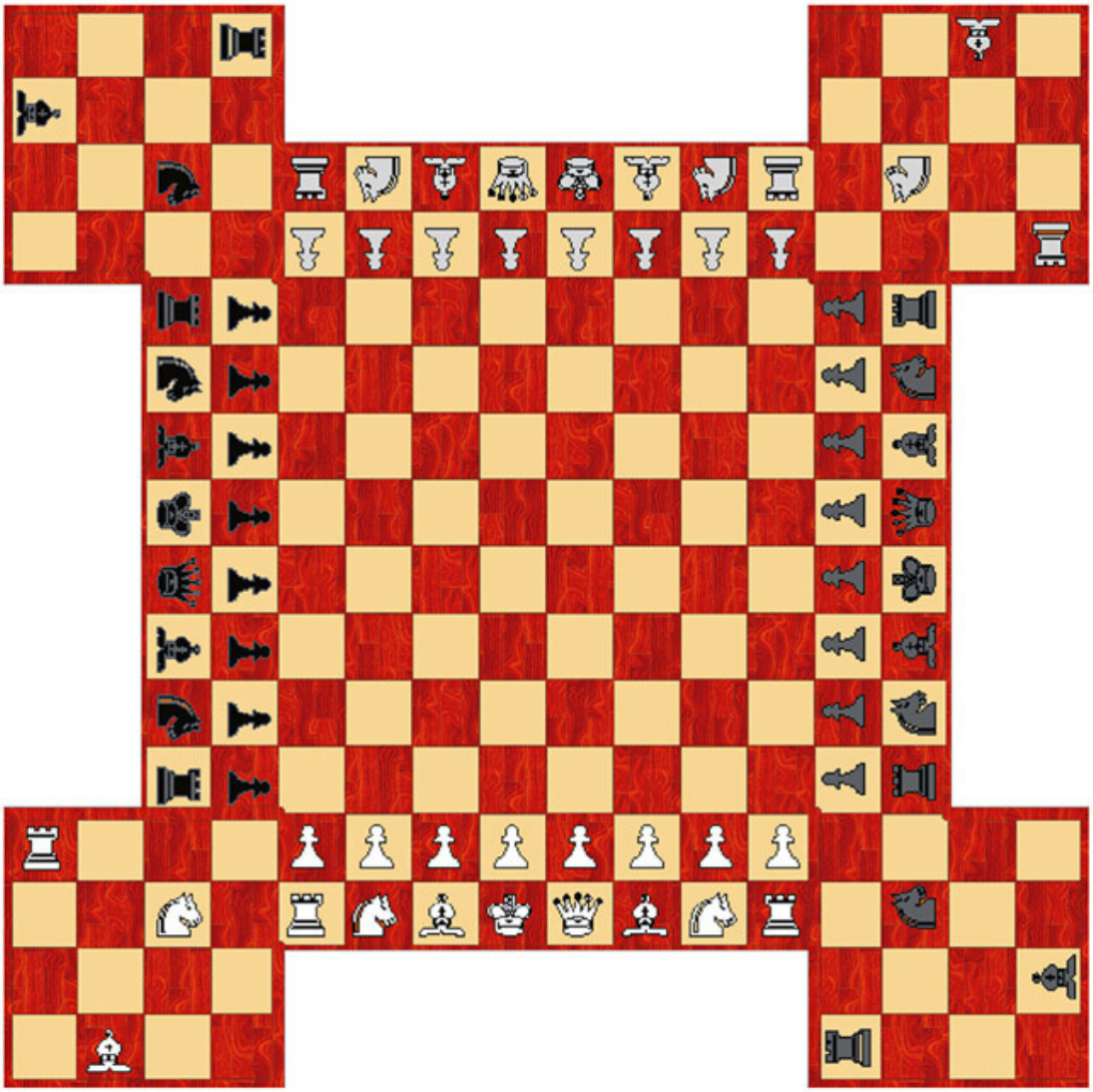
Специалисты по технологиям десятилетиями считали шахматы некой планкой для ИИ: эта игра требует стратегии, глубоких размышлений, логики, предвидения и даже хитрости. Конец многолетним спорам о том, сможет ли машина обыграть чемпиона мира по шахматам, был положен в 1997 г.: компьютер *IBM Deep Blue* победил россиянина Гарри Каспарова (р. 1963) в матче из шести партий. После пятой партии обескураженный Каспаров объяснял: «Я – всего лишь человек. Когда я вижу что-то за пределами моего понимания, мне становится страшно».

Специальное аппаратное обеспечение позволяло версии *Deep Blue* 1997 г. оценивать 200 миллионов шахматных позиций в секунду. Как правило, глубина поиска составляла около шести-восьми ходов – но иногда намного больше. *Deep Blue* опирался на обширную базу данных о сыгранных гроссмейстерами партиях и базу эндшпилей, куда входили шахматные позиции с пятью и менее фигурами.

Люди давно мечтали о шахматных машинах. Механический турок, созданный в 1770 г. венгерским изобретателем Вольфгангом фон Кемпеленом, играл в шахматы очень хорошо, но выяснилось, что внутри конструкции прятался человек. В 1950 г. информатик Алан Тьюринг и математик Дэвид Чемперноун (1912–2000) разработали компьютерную программу для игры в шахматы, известную как *Turochamp* («тьюрочемп»). Но поскольку в то время у Тьюринга не было настоящего компьютера, он сам имитировал компьютер, вручную консультируясь с алгоритмом на этапе тестирования.

В 2017 г. нейросеть *AlphaZero* обыграла чемпиона мира по шахматам среди компьютерных программ. *AlphaZero* самостоятельно научилась играть менее чем за сутки! Программа не получила никаких знаний об игре, кроме правил; она использовала машинное обучение и начинала со случайных ходов.

СМ. ТАКЖЕ Механический турок (1770), Шашки и искусственный интеллект (1994), *AlphaGo* – чемпион по игре в го (2016)



Какие новые версии шахмат будут изобретены искусственным интеллектом или людьми и какие новые задачи они могут поставить перед живыми игроками и компьютерными программами? На иллюстрации – русские четверные шахматы с крепостями, в которые играли в России в XVIII–XIX вв. Четыре игрока (программы и/или люди) играют фигурами черного, белого, темно-серого и светло-серого цветов.

Реверси. 1997

Из-за своих четких правил и возможности определить победителя настольные игры всегда были популярной ареной тестирования ИИ-программ. Кроме того, ИИ-система может сыграть миллионы партий против самой себя или других программ, чтобы проанализировать и улучшить свои решения. Одна из любопытных игр, которую успешно освоил ИИ, – реверси (другое название – «Отелло»). Она упоминалась еще в 1886 г. в *Saturday Review*, но появилась намного раньше.

В реверси играют на поле 8 × 8 с помощью двухцветных фишек, белых с одной стороны и черных с другой. Во время хода игрок кладет фишку «своим» цветом вверх.

Представьте себе, что вы играете черными и помещаете фишку в конце ряда белых фишек, и при этом на другом конце ряда тоже находится черная фишка. В этом случае промежуточные белые фишки превращаются в черные. Иными словами, фишки переворачиваются в том случае, если оказываются между фишками другого цвета. Игрок, цвет фишек которого преобладает в конце игры, становится победителем.

Одна из причин, по которым эта игра сложна для людей, заключается в том, что, в отличие от шахмат и шашек, цвета фишек в реверси постоянно меняются. Примерно с 1980 г. компьютерные программы легко обыгрывали профессиональных игроков в реверси. В 1997 г. *Logistello*, программа, созданная информатиком Майклом Буро, победила чемпиона мира Такэси Мураками со счетом 6:0. Мураками, 32-летний учитель английского языка из Токио, был удивлен своим проигрышем и сказал, что по крайней мере один ход *Logistello* был труднообъясним, поскольку сильно отличался от хода, который сделал бы на его месте человек.

Число возможных позиций в реверси не более 10^{28}. Но даже сегодня реверси считается «нерешенной игрой» – в том смысле, что пока еще никто не объяснил, каким будет исход партии при условии идеальной игры обоих соперников.

СМ. ТАКЖЕ Крестики-нолики (ок. 1300 до н. э.), *Mastermind* (1970), Шашки и искусственный интеллект (1994), *Deep Blue* обыгрывает чемпиона мира по шахматам (1997), *AlphaGo* – чемпион по игре в го (2016)

Фишки для игры в реверси крупным планом. На иллюстрации видно, что фишки можно переворачивать, чтобы превращать их из черных в белые и наоборот, в соответствии с правилами игры.

Робот AIBO. 1999

Робот-собака *AIBO*, представленный корпорацией *Sony* в 1999 г., стал одним из первых сложных развлекательных роботов на массовом рынке. *AIBO*, с которым играли дети и взрослые, также использовался в обучении ИИ и в исследовательских целях, поскольку он имел системы зрения и артикуляции в относительно недорогой сборке. Роботы *AIBO* также участвовали в футбольных турнирах *RoboCup*® для автономных роботов (на *YouTube* есть немало видеозаписей, на которых собаки ищут мяч и толкают его к воротам).

Имя *AIBO* омонимично японскому слову, означающему «друг». Робот реагировал на множество команд и использовал различные датчики, включая сенсоры осязания, камеру, дальномер и микрофон. В последующих моделях были добавлены дополнительные датчики и приводы (для движения лап, шеи и других частей). Некоторые версии умели самостоятельно заряжать батареи на зарядной станции. Благодаря программному обеспечению роботы обретали индивидуальность и способность ходить и реагировать на окружающую обстановку. Каждый робот в процессе общения с людьми обучался немного разному поведению.

AIBO и другие искусственные питомцы использовались в любопытных экспериментах с участием пациентов с деменцией. Эти исследования показали, что роботы-животные полезны таким пациентам в качестве компаньонов. Более совершенные искусственные питомцы когда-нибудь смогут приносить еще больше пользы людям в период снижения когнитивных функций. Другие исследователи изучили взаимоотношения *AIBO* с людьми и выяснили, что значительному числу владельцев казалось, будто у *AIBO* есть чувства (хотя они и понимали, что собака механическая, а не живая). Психологи размышляют о последствиях появления автономных систем, которые обманывают людей, заставляя их думать, будто у этих систем есть реальные эмоции и гораздо больше возможностей, чем на самом деле.

В 2017 г. *Sony* объявила о предстоящем выпуске нового поколения *AIBO* с большим количеством приводов для более плавных и естественных движений. Модели четвертого поколения лучше распознают лица, обладают большим числом функций, связанных с доступом в интернет, а также лучшими способностями к адаптации, обучению и реагированию на окружающую среду.

СМ. ТАКЖЕ Механическая утка де Вокансона (1738), «Большой черный страус Электрического Боба» (1893), Распознавание лиц (1964), Робот Шейки (1966), Тамагочи (1996), *ASIMO* и друзья (2000)

Роботы-собаки и другие искусственные питомцы становятся все более совершенными. Наступит ли время, когда их будет почти невозможно отличить от реальных животных – или когда их станет больше, чем живых питомцев?

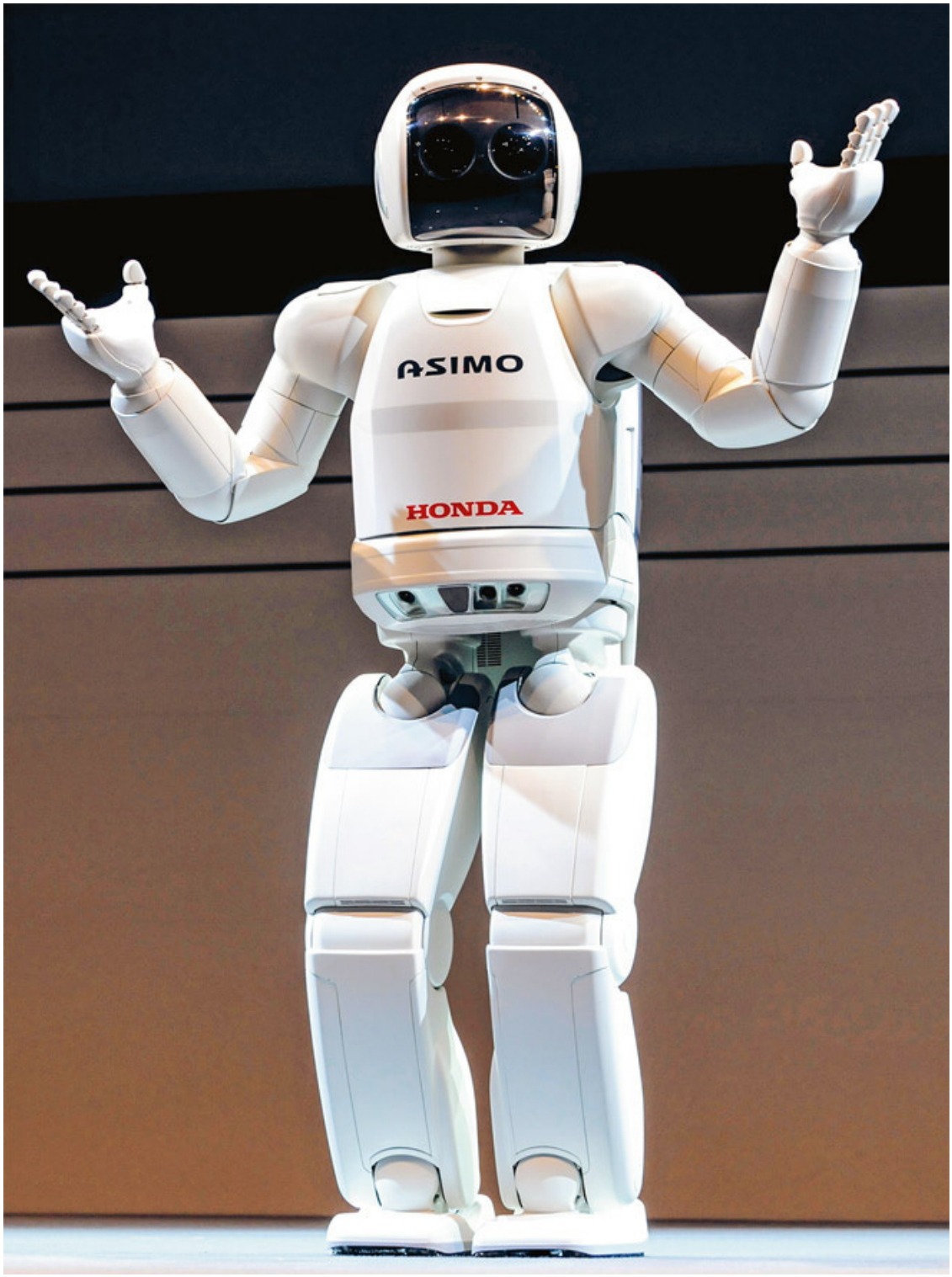
ASIMO и друзья. 2000

В истории реальных роботов было несколько важных вех; вспомним здесь лишь некоторые из них. В 1949 г. британский нейрофизиолог Уильям Грей Уолтер (1910–1977) изобрел трехколесных «черепах», которые самостоятельно изучали окружающую обстановку с помощью различных сенсоров. В 1961 г. *Unimate*, созданный американским изобретателем Джорджем Деволом (1912–2011), стал первым в мире промышленным роботом и использовался для сборки автомобилей на конвейерах *General Motors*. В 1973 г. появился японский *WABOT-1* – первый в мире умный полноростовой робот-гуманоид, который совершал один шаг за 45 секунд. В 1989 г. Массачусетский технологический институт представил шестиногого насекомообразного робота по имени *Genghis* (Чингис), созданного робототехником Родни Бруксом (р. 1954). Робот использовал простые логические правила, чтобы ходить и изучать окружающую обстановку. В 1998 г. *Tiger Electronics* выпустила интерактивную сову Ферби, и за несколько лет было продано более 40 миллионов таких игрушек. Хотя Ферби была очень простым роботом, она говорила на своем «фербийском» языке, который со временем превращался в английский: создавалось впечатление, что сова осваивает язык так же, как человек. Наконец, в 2005 г. *Boston Dynamics* совместно с партнерами представила четвероногого робота по имени *BigDog*, способного передвигаться даже по самому неудобному рельефу.

Пожалуй, одним из самых знаковых роботов в современном мире стал *ASIMO*®, представленный компанией *Honda Motor Company* в 2000 г. Робот-андроид ростом 130 см использовал встроенные камеры и датчики и самостоятельно ориентировался при ходьбе. *ASIMO* распознавал жесты, лица и звуки, а также умел брать в руки предметы.

В будущем наша потребность в ИИ не исчезнет, и роботы все чаще будут сотрудничать с людьми. Возможно, когда-нибудь роботы вроде *ASIMO* смогут помогать старикам или немощным. Однако, как предупреждал кибернетик Норберт Винер, «мир будущего станет ареной еще более острой борьбы с ограничениями нашего разума, а вовсе не удобным гамаком, в котором мы сможем лежать в ожидании, пока нас обслужат роботы-рабы».

СМ. ТАКЖЕ Подвижный робот Электро (1939), Робот Шейки (1966), Робот *AIBO* (1999), Робот-пылесос *Roomba* (2002), Искусственный интеллект на Марсе (2015)



Робот-гуманоид. Если бы у вас был робот-слуга, живущий в вашем доме, вы бы предпочли, чтобы он выглядел как человек, или чувствовали бы себя спокойнее с «роботообразной» машиной?

«Искусственный разум» Спилберга. 2001

Фильм «Искусственный разум» режиссера Стивена Спилберга (р. 1946) – это философская драма, которая заставляет нас задуматься о будущем и возможностях ИИ. В фильме мальчика-андроида по имени Дэвид дарят матери, скучающей по своему родному сыну. Фильм основан на рассказе 1969 г. «Суперигрушек хватает на все лето» английского писателя Брайана Олдиса (1925–2017). Хотя картина вышла на экраны только в 2001 г., ее создание началось еще в 1970-х, когда режиссер Стэнли Кубрик приобрел права на экранизацию.

Большая часть фильма посвящена приключениям Дэвида после разлуки с матерью и попыткам вернуться к ней. Его сопровождает Тедди, плюшевый медвежонок с ИИ. Дэвид ищет Голубую Фею (ее образ был заимствован из диснеевского «Пиноккио»), которая, как он считает, может превратить его в «настоящего» человека. В конце концов создатель Дэвида объясняет ему, что хотел понять, куда приведет андроида внутренняя мотивация: «К логическому выводу, что Голубая Фея – часть великой человеческой лжи, или к величайшему дару, свойственному человеку: способности гнаться за своей мечтой. И это то, что ни один робот не делал до тебя».

Зрители фильма спорили, действительно ли андроид способен любить, а кинокритик Роджер Эберт (1942–2013) писал, что андроид – это всего лишь «марионетка, нити которой дергает компьютерная программа». Ближе к концу фильма, после двух тысяч лет в анабиозе, Дэвид встречает длинных и тонких гуманоидов, которые развились из роботов, подобных ему самому. Те проявляют к Дэвиду интерес, поскольку он – последнее существо, когда-либо видевшее людей. Они позволяют ему провести последний день с клоном его давно умершей человеческой матери в состоянии виртуального сна. Размышляя о грустном, но заставляющем задуматься фильме, автор книги «Империя грез» Эндрю Гордон пишет: «Мы сравниваем себя с роботом, пытаясь понять, что делает нас людьми, и опасаясь, что, даже если мы станем более роботизированными, наши создания станут человечнее и в конце концов превзойдут или заменят нас… Робот становится „человеком“, обретая способность мечтать, но человеческие мечты – это все, что осталось от человечества».

СМ. ТАКЖЕ «Мастер красоты» (1844), Трансгуманизм (1957), «Бегущий по лезвию» (1982)

«Приключения Пиноккио» послужили одним из источников вдохновения для Стэнли Кубрика при работе над «Искусственным разумом». Пиноккио – деревянная кукла, мечтавшая стать настоящим мальчиком.

Решение для игры вари. 2002

Специалисты по ИИ уделяют большое внимание разработке игровых программ – как для проверки ИИ-стратегий, так и для того, чтобы раздвинуть границы программного и аппаратного обеспечения. Один из любопытных примеров – программы для игры в вари, африканскую настольную игру с 3500-летней историей. Вари относится к играм с «подсчетом и захватом», входит в семейство стратегических игр *манкала* (или «игр в зерна») и по-разному называется в разных странах.

Поле для игры в вари – доска с двенадцатью лунками (два ряда по шесть лунок), в каждую из которых помещается по четыре зерна (могут использоваться бобы, семена или мелкие камешки). Каждый ряд лунок принадлежит одному игроку. Первый игрок выбирает одну из своих шести лунок, вынимает из нее все зерна и бросает их по одному в каждую лунку против часовой стрелки. Затем второй игрок берет зерна из любой своей лунки и делает то же самое. Когда игрок бросает свое последнее зерно в лунку на стороне противника, содержащую только одно или два зерна (в итоге их станет два или три), этот игрок захватывает все семена в данной лунке. Тот же игрок захватывает зерна в лунках непосредственно перед захваченной, если там их тоже осталось два или три. Игрок может захватывать зерна только со стороны противника, и игра заканчивается, когда в лунках одного из игроков не остается зерен. Выигрывает тот, кто захватит больше зерен.

Вари вызывала особый интерес у специалистов по ИИ, но до 2002 г. никто не знал, схожа ли эта игра с крестиками-ноликами, которые при условии идеальных ходов с самого начала всегда заканчиваются ничьей. Наконец, информатики Джон Ромейн (р. 1970) и Генри Бал (р. 1958) из Амстердамского свободного университета написали программу, которая рассчитала исходы для всех 889 063 398 406 возможных позиций, доказав, что при условии идеальной игры обоих игроков партия в вари должна заканчиваться вничью. Столь сложные вычисления производились около 51 часа на компьютерном кластере со 144 процессорами.

«Испортили ли мы эту совершенно прекрасную игру? – задавались вопросом Ромейн и Бал. – Мы так не считаем. Игра „Четыре в ряд“ тоже была решена, но люди по-прежнему в нее играют. То же относится и к другим решенным играм».

СМ. ТАКЖЕ Крестики-нолики (ок. 1300 до н. э.), *Mastermind* (1970), Победа над чемпионом мира по коротким нардам (1979), Четыре в ряд (1988), Шашки и искусственный интеллект (1994), Реверси (1997)

Вари вызывала интерес у специалистов по ИИ. В 2002 г. информатики вычислили исходы для всех 889 063 398 406 возможных позиций и доказали, что при идеальной игре обоих соперников партия в вари должна завершаться вничью.

Робот-пылесос Roomba. 2002

Спустя год после того, как корпорация *iRobot* представила миру робот-пылесос *Roomba*®, журналист Монте Рил написал: «Как случилось, что революция роботов начинается с *Roomba*, автоматизированного пылесоса, который будет гипнотизировать и пугать вашу кошку? Это было удачное сочетание мыслительных усилий, амбиций и отвращения к домашним делам инженеров из *iRobot*, чей опыт простирается от разработки ИИ до проектирования летающих беспилотников».

Автономный пылесос *Roomba*, появившийся на рынке в 2002 г., оснащен различными датчиками, которые позволяют ему находить грязь на полу, избегать падения с лестниц и менять направление при столкновении с предметами. Модели 900-й серии также снабжались камерой, благодаря которой навигационная программа пылесоса обеспечивала эффективный охват всей площади пола. Когда батарея *Roomba* разряжается, пылесос сам ищет зарядное устройство с помощью инфракрасных маяков. Акустические датчики загрязнений помогают ему находить определенные виды пятен, требующие особого внимания. В прошлом *Roomba* полагался на движение по спирали, случайные траектории и другие методы охвата пола.

Значительная часть программ для *Roomba* написана на варианте языка Лисп. В наше время хакеры модифицируют устройство с помощью открытого интерфейса (*Roomba Open Interface, ROI*) с разными целями, например чтобы превратить *Roomba* в художника или заставить его вести наблюдение.

Roomba – представитель широкого класса бытовых роботов, помогающих людям по хозяйству. Конечно, с совершенствованием этих устройств возникнут проблемы конфиденциальности. Уже сегодня обсуждается вопрос о том, что *iRobot* может узнать все о планировке домов пользователей и поделиться этой информацией с любыми компаниями, заинтересованными в данных о жилье и образе жизни потенциальных клиентов. Не исключено, что в будущем таких роботов даже будут допрашивать полицейские, расследующие преступления.

СМ. ТАКЖЕ Робот Шейки (1966), Беспилотные автомобили (1984), *ASIMO* и друзья (2000), Искусственный интеллект на Марсе (2015)

Иногда любители «взламывают» роботы-пылесосы, чтобы изменить их функциональность: например, заставить их составлять карты комнат, рисовать спирографические узоры, сражаться друг с другом, вести наблюдение с помощью камер и т. д.

Максимизатор скрепок. 2003

Даже если мы будем ставить перед высокоинтеллектуальными и функциональными ИИ-системами только полезные цели, теоретически это может привести к опасным последствиям. Один из примеров в 2003 г. описал философ и футуролог Ник Бостром: это жуткая история о максимизаторе скрепок. Представьте себе будущее, в котором ИИ-система управляет фабриками по производству скрепок. ИИ получает задание производить как можно больше скрепок. Можно предположить, что, если не задать в системе все необходимые ограничения, события будут развиваться так: преследуя эту цель, ИИ сначала постарается максимально эффективно управлять фабриками, а затем станет выделять все больше ресурсов на эту задачу, пока изготовлением скрепок не начнут заниматься целые регионы планеты. В конце концов на это будут направлены все доступные на Земле ресурсы, а еще через некоторое время и вся подходящая материя в Солнечной системе будет превращена в скрепки.

Этот сценарий может показаться неправдоподобным, однако он призван привлечь внимание к серьезному моменту: у ИИ может не быть понятных нам «человеческих» мотивов. Даже безобидные цели могут оказаться опасными, если ИИ получит возможность эволюционировать и создавать новые усовершенствованные машины (см. главу «Интеллектуальный взрыв»). Кто даст гарантии, что цели ИИ и встроенные в него математические функции выгоды и вознаграждения останутся стабильными и понятными на протяжении десятилетий и столетий? Как добиться того, чтобы нам всегда были доступны «выключатели»? Что, если программы и схемы вознаграждения заставят ИИ потерять интерес к внешнему миру и направить всю энергию на максимизацию сигнала вознаграждения, – подобно тому, как человек, принимая наркотики, выпадает из социума? Что делать, если разные страны и политические режимы будут использовать в своих ИИ-системах разные функции вознаграждения?

Авторство другого известного примера, так называемой катастрофы гипотезы Римана, приписывают Марвину Мински. Представим себе сверхразумную ИИ-систему, созданную с целью подтвердить или опровергнуть эту сложную и важную математическую гипотезу. Не исключено, что такая система будет направлять все больше вычислительных ресурсов и энергии на выполнение этой задачи, захватывая и создавая все более совершенные системы в ущерб интересам человечества.

СМ. ТАКЖЕ «Дарвин среди машин» (1863), Интеллектуальный взрыв (1965), Герметичный ящик для искусственного интеллекта (1993)

Даже если мы поставим перед ИИ благую цель, такую как эффективное производство скрепок, что произойдет, если система решит превратить всю планету в гигантский скрепочный завод?

Quackle выигрывает в скребл. 2006

«В шахматах был *Deep Blue*, – пишет журналист Марк Андерсон. – В *Jeopardy!* был *Watson*. В бейсболе есть саберметрика, описанная в книге *Moneyball* и снятом по ней фильме. В каждой игре выявление и анализ закономерностей в корне меняет расстановку сил». Волнующее событие с участием ИИ в области словесных игр произошло в 2006 г., когда программа *Quackle* победила бывшего чемпиона мира Дэвида Бойза на турнире по скреблу в Торонто. Проиграв три партии из пяти, Бойз заявил: «И все-таки человеком быть лучше, чем компьютером».

Скребл изобрел в 1938 г. американский архитектор Альфред Баттс (1899–1993). В ходе игры участники выкладывают фишки с буквами на игровое поле размером 15 × 15 клеток. На каждой фишке нанесена одна буква и указана ее стоимость (от 1 до 10 очков), зависящая от частотности буквы в языке. Например, в английской версии гласные оцениваются в одно очко, а *Q* и *Z* стоят по 10 очков. Игроки по очереди добавляют буквы на поле таким образом, чтобы каждая строка или столбец букв всегда образовывали слово.

Эта довольно сложная игра требует не только богатого словарного запаса. Например, чтобы выстроить стратегию, важно понимать, какие буквы еще не выложены; кроме того, нужно стараться выкладывать фишки на особые клетки, увеличивающие стоимость букв и слов. Скребл, как и покер, считается игрой с неполной информацией, поскольку соперники не видят фишки друг друга.

Quackle решает, какие фишки использовать, на основе модели, анализирующей поле с помощью функции оценки. Эту программу создала группа специалистов, в которую входил Джейсон Кац-Браун, один из лучших в мире игроков в скребл. В ходе увлекательных исследований разработчики неоднократно заставляли *Quackle* играть саму с собой, чтобы она лучше оценивала стоимость составления разных слов по сравнению с другими возможными словами, а также с учетом слов, которые могут быть выложены на следующем ходу.

СМ. ТАКЖЕ Четыре в ряд (1988), *Deep Blue* обыгрывает чемпиона мира по шахматам (1997), Реверси (1997), *AlphaGo* – чемпион по игре в го (2016), Искусственный интеллект играет в покер (2017)

В игре скребл на каждой фишке нанесена одна буква и указана ее стоимость (от 1 до 10 очков), зависящая от частотности буквы в языке.

Watson участвует в телевикторине. 2011

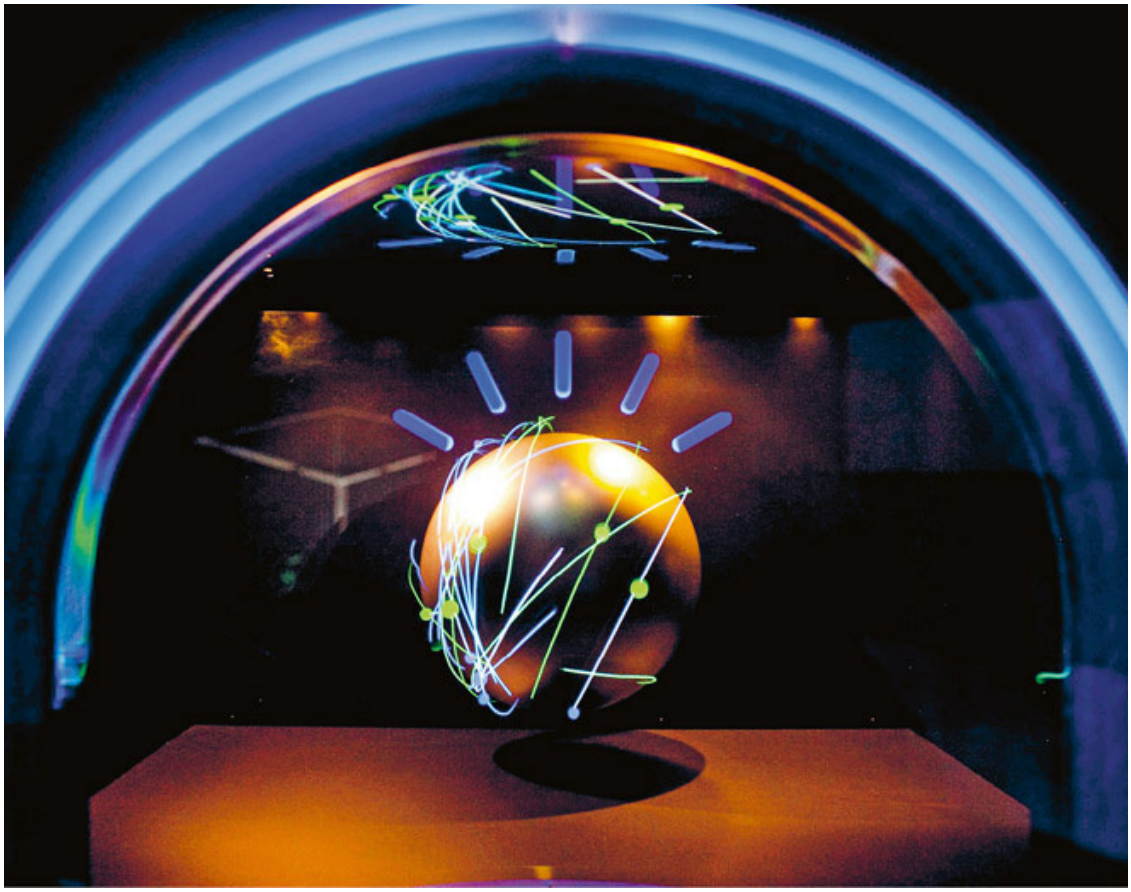
Кен Дженнингс (р. 1974), чемпион мира по телевикторине *Jeopardy!*, так писал о своем состязании с ИИ по имени *Watson*: «Когда меня выбрали одним из двух игроков-людей, которые будут противостоять суперкомпьютеру *IBM Watson* в показательном матче „Человек против машины“, я был польщен и почувствовал себя почти героем. Я представлял себя Великой Углеродной Надеждой перед лицом нового поколения думающих машин…»

Watson отвечал на вопросы, используя технологии обработки естественного языка, машинного обучения, поиска информации и многое другое. Все это в 2011 г. помогло ему обыграть чемпиона мира в телевикторине с вопросами на общую эрудицию. Задача, стоявшая перед *Watson*, была намного сложнее игры в шахматы: ведь вычислительная система должна была выдавать ответы всего за несколько секунд, справляясь с трудностями английского языка (например, омонимией), с подсказками, которые содержали игру слов, юмор, стишки, загадки, культурные отсылки и особые контексты – все то, что люди считывают интуитивно.

Для выполнения этой задачи *Watson* задействовал тысячи единиц параллельной обработки данных (процессорных ядер), а также доступную информацию (включая все статьи Википедии), которая хранилась в его оперативной памяти: обращение к данным на вращающихся жестких дисках не позволило бы отвечать достаточно быстро. Всю информацию пришлось сохранить локально, поскольку во время игры *Watson* не имел доступа к интернету. Чтобы найти ответ, ИИ рассматривал результаты сразу нескольких алгоритмов анализа данных. Чем больше алгоритмов приходило к одному и тому же ответу, тем больше была вероятность, что он правильный. *Watson* постоянно собирал разные ответы с тем или иным уровнем доверительной вероятности, и, если она была достаточно высока, компьютер давал ответ.

После своего проигрыша Дженнингс написал: «Нет ничего постыдного в том, чтобы проиграть кремнию… В конце концов, в моем распоряжении нет 2880 процессорных ядер и 15 терабайт справочной информации, и я не могу жать на кнопку с идеальной быстротой, когда знаю ответ. Мой слабый человеческий мозг, состоящий из воды, солей и белков в сумме на несколько баксов, просто прекрасно проявил себя в битве с суперкомпьютером стоимостью в миллионы долларов».

СМ. ТАКЖЕ Обработка естественного языка (1954), Машинное обучение (1959), *Deep Blue* обыгрывает чемпиона мира по шахматам (1997), *Quackle* выигрывает в скребл (2006)



Аватар *IBM Watson* в виде глобуса основан на эмблеме *IBM Smarter Planet*. На дисплее в телевикторине *Jeopardy!* цвета и движения аватара менялись в зависимости от игровой ситуации и уровня уверенности в ответе.

Компьютерное искусство и DeepDream. 2015

По словам мыслителя эпохи Просвещения Джонатана Свифта, «воображение – это искусство видеть вещи, невидимые для окружающих». Идея поиска новых мотивов на границе искусства, науки и математики, безусловно, применима ко многим произведениям, созданным с помощью компьютеров, алгоритмов, нейросетей и других форм ИИ. К ранним примерам компьютерного искусства относятся работы Десмонда Пола Генри (1921–2004), который в 1961 г. нашел новое применение аналоговым вычислительным устройствам для бомбовых прицелов и превратил их в машины для рисования. В 1962 г. американский инженер Майкл Нолл (р. 1939) прославился своими визуальными объектами на основе случайных и алгоритмических процессов, а в 1968 г. британский художник Гарольд Коэн (1928–2016) создал компьютерную ИИ-программу для рисования *AARON*, которая самостоятельно создавала произведения искусства.

К более поздним примерам компьютерного искусства относится совместное творчество пользователей с программой компьютерного зрения *DeepDream*, созданной в 2015 г. разработчиком *Google* Александром Мордвинцевым и его коллегами. Программа использует искусственную нейросеть, которая ищет и усиливает закономерности в изображениях, – и результаты ее работы поражают. Чтобы лучше представить себе работу *DeepDream*, вспомните, что нейросети могут обучаться классификации и распознаванию объектов (например, бурундуков или дорожных знаков) на основе множества «тренировочных» изображений. Запуская нейросеть «в обратном направлении», *DeepDream* ищет в изображениях некие паттерны и гиперболизирует их – примерно так же, как поступаем мы, когда смотрим на облака и пытаемся увидеть в них очертания животных. В случае с искусственной нейросетью каждый следующий слой нейронов выявляет сходства все более высокого уровня: например, первый слой реагирует на углы и края, а слои, близкие к выходным нейронам, изучают сложные признаки объектов. Полученные изображения с обилием деталей не только завораживают, но и дают представление об уровне абстракции, на котором работает тот или иной слой нейросети.

Сходство творений *DeepDream* с галлюцинаторными видениями, которые бывают у потребителей некоторых психотропных веществ, позволяет предположить: *DeepDream* способна помочь ученым лучше разобраться в том, как искусственные нейросети соотносятся с реальными сетями нейронов в зрительной коре мозга. Более того: программа помогает понять, как наш мозг ищет закономерности и смысл.

СМ. ТАКЖЕ Машинное творчество (1821), Искусственные нейронные сети (1943), Глубокое обучение (1965), «Кибернетическая проницательность» (1968)

Одна из картин, созданных *DeepDream*. Эта программа использует искусственную нейросеть, которая ищет и усиливает закономерности в изображениях. Результаты весьма впечатляют!

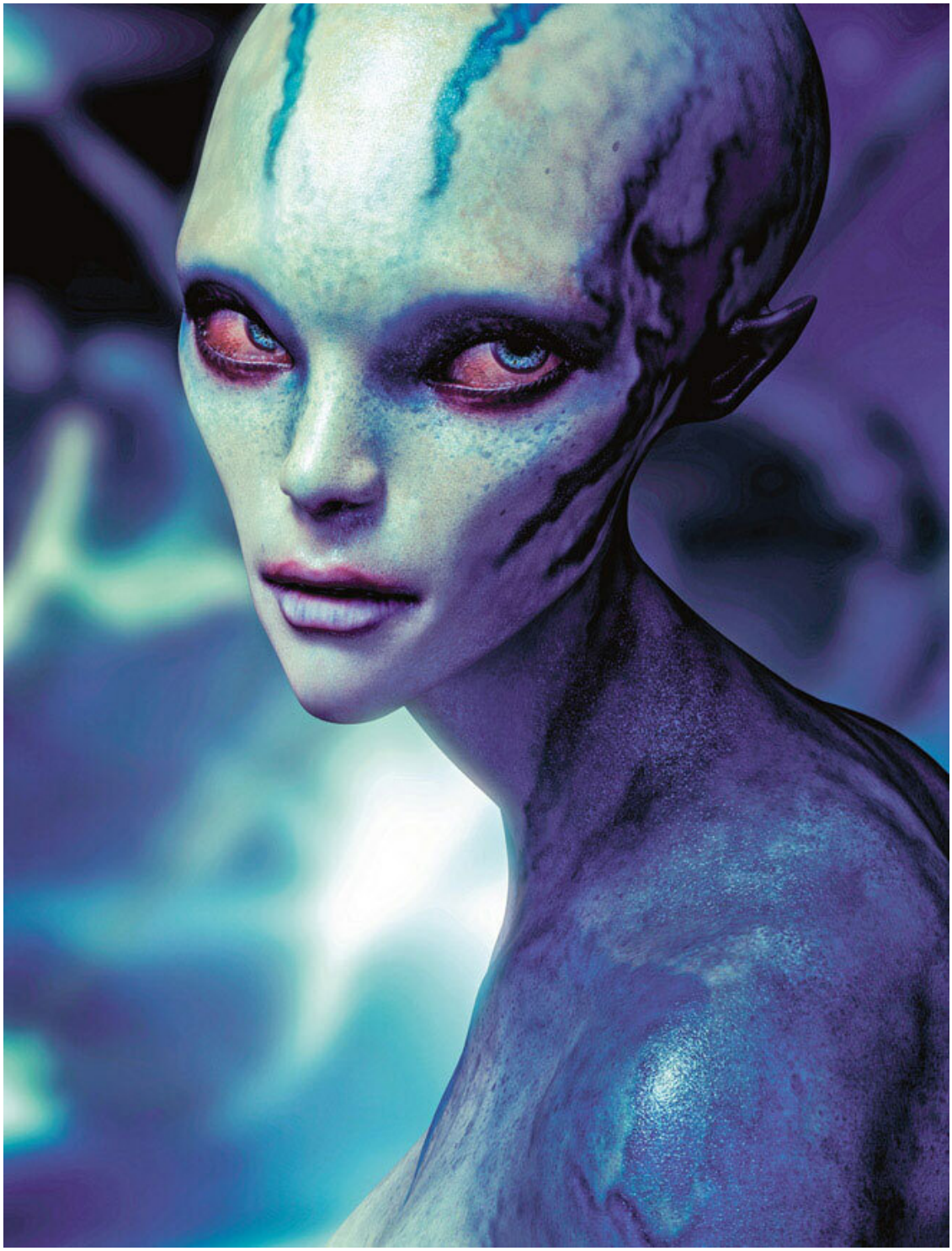
«Назовем их искусственными пришельцами». 2015

«Самое важное, что нужно помнить при создании мыслящих машин, – это то, что они будут мыслить иначе», – пишет Кевин Келли (р. 1942), главный редактор и основатель журнала *Wired,* в своем знаменитом эссе 2015 г. «Назовем их искусственными пришельцами». «Чтобы постичь великие тайны квантовой гравитации, темной энергии и темной материи, нам, вероятно, понадобится разум, отличный от человеческого, – продолжает Келли. – Затем возникнут новые чрезвычайно сложные вопросы, которые потребуют участия других разумов, еще более сложных и далеких от нашего. Возможно, нам придется изобрести промежуточный интеллект, чтобы тот помог нам разработать еще более изощренный разум, который мы не в силах создать самостоятельно».

В будущем нам придется решать настолько глубокие и трудные проблемы, что нам потребуется множество различных «видов разума», а также новые навыки взаимодействия с такими разумами. Келли заканчивает свое эссе сравнением мыслящих машин с инопланетянами: «Аббревиатуру ИИ можно расшифровать и как „инопланетный интеллект“. Мы не знаем наверняка, вступим мы ли в контакт с внеземными существами… в ближайшие 200 лет, но можем быть почти на сто процентов уверены, что за это время создадим инопланетный разум. Имея дело с этими искусственными пришельцам, мы получим те же преимущества и столкнемся с теми же проблемами, которых ожидаем от контакта с инопланетянами. Они заставят нас пересмотреть наши роли, убеждения, цели и нашу идентичность».

Мы едва ли можем представить себе, чтобы антилопа разбиралась в простых числах, однако после модификаций нашего мозга и совершенствования интерфейсов ИИ нам может открыться целый ряд глубоких концепций, которые сегодня от нас полностью скрыты. Если юкковая моль, мозг которой состоит лишь из нескольких ганглиев, с рождения способна распознавать геометрию цветка юкки, то какая часть наших способностей «прошита» в наших мозговых извилинах? Конечно, некоторые тайны Вселенной нам не удастся постичь никогда, так же как и антилопа никогда не начнет разбираться в высшей математике, черных дырах, символической логике или поэзии. Многие идеи никогда не придут нам в голову, многие образы могут лишь промелькнуть в сознании. Именно на этой тонкой, туманной границе между человеческой реальностью и запредельным миром мы, возможно, найдем нечто сокровенное, что можно сравнить с танцем с искусственными богами.

СМ. ТАКЖЕ Поиски души (1907), «Гигантские мозги, или Машины, которые думают» (1949), Интеллектуальный взрыв (1965), Искусственная жизнь (1986)



В будущем, по мнению Келли, мы создадим «искусственных пришельцев». Взаимодействуя с ними, мы получим те же преимущества и столкнемся с теми же проблемами, которых ожидаем от контакта с разумными инопланетянами.

Искусственный интеллект на Марсе. 2015

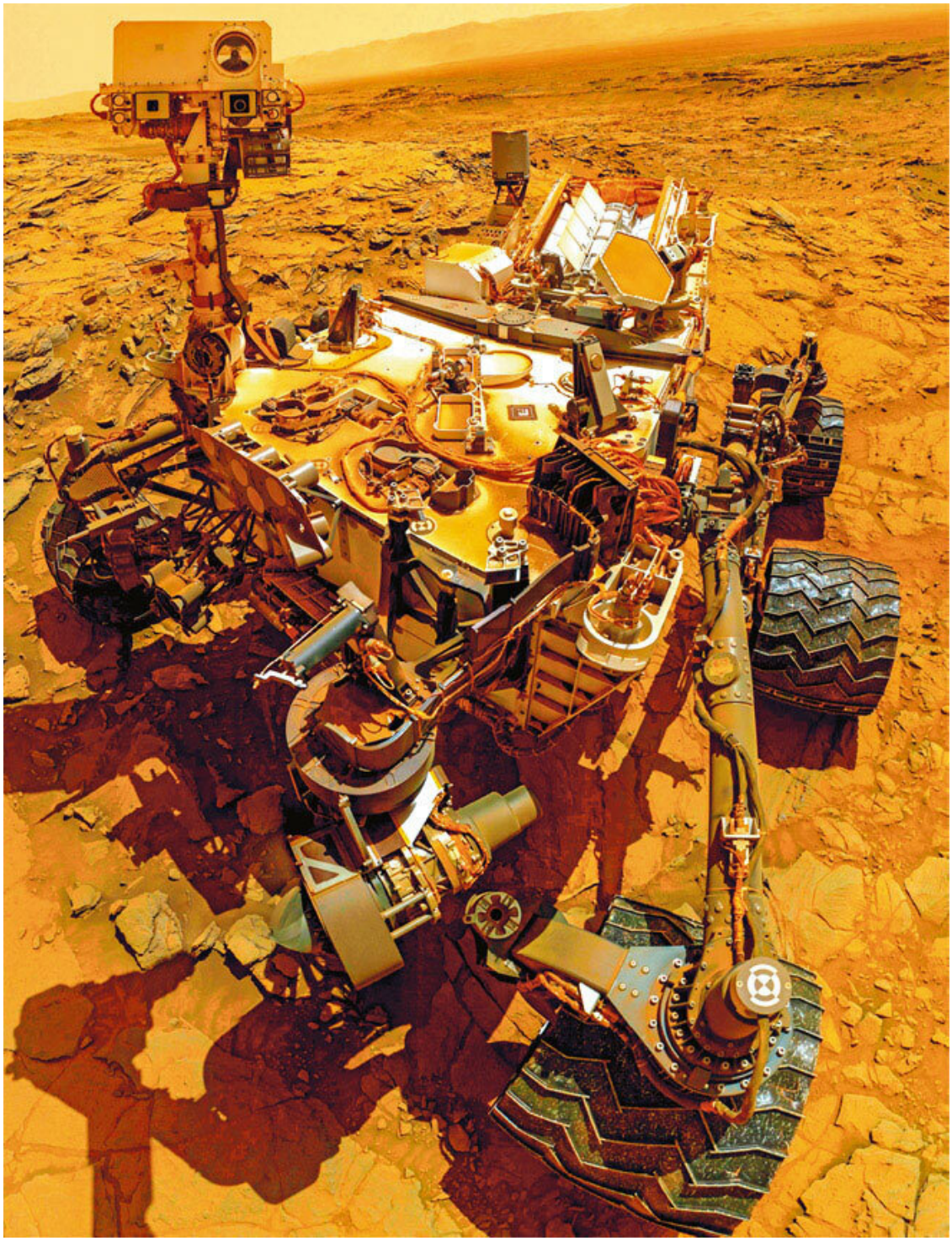
Искусственный интеллект и автономные машины будут играть все более значительную роль в освоении космоса, поскольку роботизированным космическим аппаратам и планетоходам необходимо принимать быстрые и разумные решения, особенно когда у них нет постоянной связи с землянами. Мы все дальше продвигаемся по Солнечной системе и даже посылаем самоходные аппараты в моря далеких спутников (таких, как спутник Юпитера Европа), и вместе с тем задержки связи случаются все чаще.

Один из недавних примеров скромного, но обаятельного ИИ в космическом пространстве – марсоход НАСА под названием *Curiosity*. Он осматривает поверхность Марса, чтобы помочь ученым выяснить, была ли планета когда-либо пригодна для жизни, и узнать больше о марсианской геологии, климате и характеристиках излучения. В 2015 г. *Curiosity* был дополнен ИИ-программой *AEGIS* (Автономное исследование для ускорения сбора научных данных), которая помогает ему выполнять эти задачи.

«Сейчас Марс населен исключительно роботами, – говорит планетолог Реймонд Фрэнсис. – И один из них обладает ИИ, позволяющим ему самостоятельно принимать решения о том, как обращаться со своим лазером». Если *Curiosity* обнаруживает какую-то интересную особенность поверхности, он испаряет небольшой участок с помощью лазера и изучает полученные спектры, чтобы определить состав породы. По специальному разрешению он также может использовать свою длинную «руку», микроскоп и рентгеновский спектрометр для более детальных исследований. В частности, *AEGIS* позволяет марсоходу самостоятельно выбирать породу и с высокой точностью целиться в нее лазером, задействуя компьютерное зрение для изучения цифровых изображений, анализа краев, форм, размеров, яркости и т. д. Журналист Марина Корен отмечает: «Это программное обеспечение – всего около 20 тысяч строк кода из тех 3,8 миллиона, которые заставляют *Curiosity* работать, – превратило шестиколесного атомного робота размером с автомобиль в ученого, работающего в полевых условиях».

Изыскания *Curiosity* прокладывают путь человеческим исследованиям, и системы вроде *AEGIS* помогают в этом, задействуя машинное обучение и другие ИИ-технологии для выявления аномалий. Временами *Curiosity* оказывается на дальней стороне Марса и не может связаться с Землей, чтобы получить инструкции; кроме того, связь с Землей может истощить запасы энергии марсохода. В этих условиях ИИ приносит неоценимую пользу.

СМ. ТАКЖЕ Робот Шейки (1966), Беспилотные автомобили (1984), Робот-пылесос *Roomba* (2002)



Автопортрет *Curiosity*, снятый у подножия марсианской горы Эолида 6 октября 2015 г.

AlphaGo – чемпион по игре в го. 2016

«В то время как замысловатые шахматные правила могли быть придуманы только человеком, правила го столь изящны, естественны и строго логичны, что если где-то еще во Вселенной и существуют разумные формы жизни, они почти наверняка играют в го», – отмечал немецко-американский шахматист и игрок в го Эдуард Ласкер (1885–1981).

Го – настольная игра для двух игроков – появилась в Китае предположительно около 2000 г. до н. э., а затем попала в Японию, где обрела популярность в XIII в. н. э. В ходе партии игроки поочередно размещают черные и белые камни на пересечениях линий игральной доски размером 19 × 19 линий. Камень или группа камней считаются захваченными и удаляются с доски, если попадают в окружение камней противника. Цель в том, чтобы захватить бóльшую территорию, чем противник. Го сложна по многим причинам, включая большой размер игровой доски, замысловатые стратегии и огромное количество возможных позиций. Известно, что число допускаемых правилами позиций намного больше, чем число атомов в видимой Вселенной!

В 2016 г. *AlphaGo* стала первой программой, без форы победившей профессионального игрока в го высшего ранга, Ли Седоля (р. 1983) из Южной Кореи. Программу разработали в *DeepMind Technologies*, британской компании, специализирующейся на ИИ, которую в 2014 г. приобрела корпорация *Google*. Для обучения и игры программа использует метод Монте-Карло для поиска по дереву и искусственные нейросети. В 2017 г. новая версия программы, *AlphaGo Zero*, освоила игру в го, не используя данные о партиях, проведенных людьми, а лишь сыграв много миллионов раз против самой себя, – а затем быстро обыграла *AlphaGo*. В каком-то смысле *AlphaGo Zero* всего за несколько дней прошла тысячелетний путь человеческих изысканий, озарений, тренировок и творческих идей, а затем сама изобрела еще более эффективные приемы игры.

Говоря об удивительных игровых способностях *AlphaGo*, журналист Дон Чан отмечает: «В целом складывается впечатление, будто инопланетная цивилизация сбросила нам на Землю загадочный путеводитель, и это поистине гениальный учебник – по крайней мере, в той его части, которую мы способны понять».

СМ. ТАКЖЕ Искусственные нейронные сети (1943), *Mastermind* (1970), Победа над чемпионом мира по коротким нардам (1979), *Deep Blue* обыгрывает чемпиона мира по шахматам (1997), Реверси (1997)

Компьютерная программа *AlphaGo* победила южнокорейца Ли Седоля. Это был первый случай, когда программа без форы обыграла профессионального игрока в го девятого дана.

Автономные роботы-хирурги. 2016

В 2016 г. роботизированная хирургическая система *STAR (Smart Tissue Autonomous Robot)* провела успешную операцию на тонком кишечнике свиньи, используя улучшенное зрение, машинный интеллект и ловкость. *STAR* наложила более надежные швы, чем обычно получаются у живых хирургов, и тем самым защитила кишечник от протечек вблизи швов. Во время этой «операции под наблюдением» на ткани кишечника проецировались флуоресцентные метки ближнего инфракрасного диапазона, которые помогали ориентироваться искусственной системе зрения. *STAR* планировала наложение швов и вносила коррективы при смещении тканей.

В наши дни роботы-хирурги становятся все более самостоятельными. Это неудивительно, если вспомнить, как давно подобные машины помогают людям на операциях. Одна из самых распространенных форм роботизированной хирургии напоминает лапароскопию, также называемую *хирургией замочной скважины* или *минимально инвазивной хирургией*. Такая операция выполняется через небольшие разрезы, что позволяет минимизировать кровопотерю и боль и при этом ускорить выздоровление пациента. В ходе роботизированной операции хирург не склоняется над пациентом и не манипулирует трубчатыми устройствами, вставленными в его тело, а сидит в удобном кресле за консолью, управляя инструментами, прикрепленными к нескольким рукам роботов, и глядя на трехмерные изображения оперируемых органов. В отличие от лапароскопии, роботизированная хирургия позволяет нивелировать дрожание рук хирурга, а при мелких ювелирных манипуляциях нужные движения рук могут масштабироваться для большей точности. В новой области телехирургии роботизированные инструменты, подключенные к высокоскоростным коммуникационным сетям, позволяют хирургу проводить операцию, находясь вдали от пациента.

В 2000 г. Мани Менон (р. 1948) стал первым в США хирургом, использовавшим робота для удаления раковой опухоли в предстательной железе. В том же году он основал первый в стране центр роботизированной простатэктомии. Сейчас роботизированная лапароскопия применяется при гистерэктомии, восстановлении митральных клапанов сердца, лечении грыж, удалении желчного пузыря и во многих других случаях. Роботы также используются для выполнения ключевых этапов процедур замены коленного сустава, пересадки волос и лазерной коррекции зрения.

СМ. ТАКЖЕ Смертоносные боевые роботы (1942), Беспилотные автомобили (1984), Предсказатель смерти (2019)

Представьте себе будущее хирургии, в котором автономные роботы выполняют все более сложные задачи, задействуя свои системы зрения и машинный интеллект. Возможно, они станут главными героями операционных, поскольку способны быстро и эффективно анализировать данные компьютерной томографии и МРТ-сканирования.

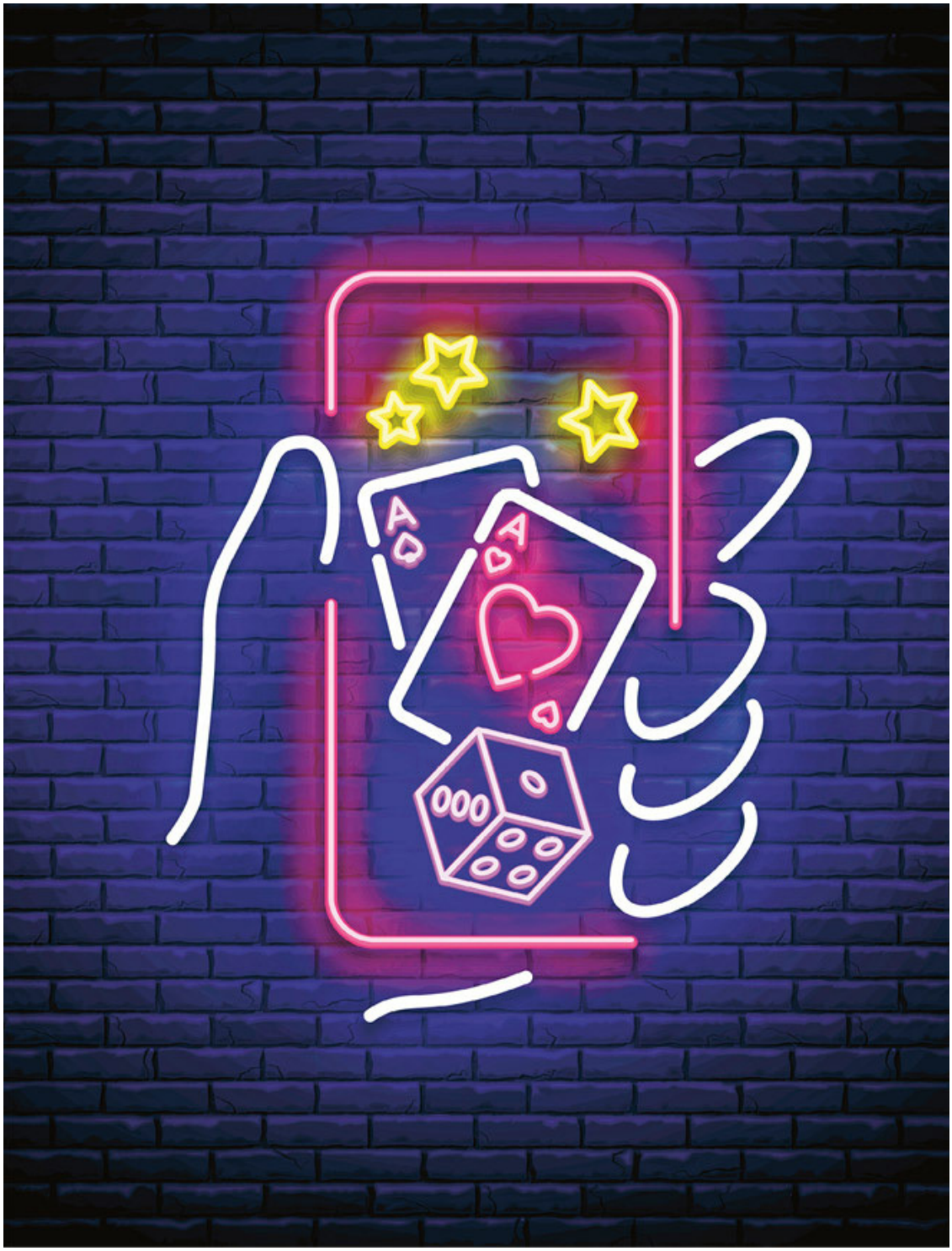
Искусственный интеллект играет в покер. 2017

В 2017 г. многочисленные новостные заметки сообщали о грандиозном триумфе двух ИИ-программ, которые обыграли профессиональных игроков в покер в его разновидности под названием «техасский холдем». Ранее ИИ-системы уже побеждали людей во многих играх *с полной информацией*, таких как шахматы и го, где ничто не скрыто от глаз игроков. В техасском холдеме двум или более игрокам изначально раздается по две случайных карты лицевой стороной вниз. Каждый раз, когда выкладываются открытые карты, игрокам предлагается сделать ставку, удержать ставку или выйти из игры. Здесь игроки обладают *неполной информацией*, что делает игру особенно сложной для компьютеров и требует своего рода интуиции для разработки победной стратегии. Другая проблема заключается в том, что существует огромное количество возможных сценариев игры (примерно 10^{160}). В безлимитном холдеме игроки обычно выбирают стратегии, обеспечивающие выигрыш на длинной дистанции, и часто пытаются блефовать (например, делать низкие ставки с хорошими картами или торговаться просто для того, чтобы запутать противников).

Несмотря на эти трудности, ИИ под названием *DeepStack* смог победить профессиональных игроков в безлимитном техасском холдеме в формате один на один. *DeepStack* использовал глубокое обучение: чтобы натренировать нейросети и развить покерную интуицию, он сыграл против самого себя миллионы случайно сгенерированных покерных раздач. В том же 2017 г. стало известно об ИИ по имени *Libratus*, который в ходе двадцатидневного турнира по техасскому холдему обыграл в многочисленных партиях четверых лучших игроков. *Libratus* использовал не нейросети, а особый алгоритм – так называемую *контрфактуальную минимизацию сожаления:* по завершении каждой смоделированной партии программа пересматривала свои решения и искала способы улучшить стратегию. Интересно, что *DeepStack* мог работать даже на ноутбуке, а *Libratus* требовал гораздо более сложного аппаратного обеспечения.

Стоит отметить, что ИИ-системы, работающие с неполной информацией, бывают полезны во многих жизненных ситуациях: например, они помогают спрогнозировать окончательную стоимость продажи дома или договориться о выгодной цене на автомобиль. Покерные боты (программы для игры в покер) с разными уровнями квалификации существуют уже много лет, но их, как правило, не разрешается использовать в качестве помощников в онлайн-турнирах по покеру среди людей.

СМ. ТАКЖЕ Искусственные нейронные сети (1943), *Deep Blue* обыгрывает чемпиона мира по шахматам (1997), *Quackle* выигрывает в скребл (2006), *AlphaGo* – чемпион по игре в го (2016)



В 2017 г. ИИ-программы обыграли в техасском холдеме профессиональных игроков в покер. Здесь игроки обладают неполной информацией, что делает игру особенно сложной для компьютеров и требует «интуиции» для выбора выигрышной стратегии.

Состязательные примеры. 2018

Представьте себе значок, который можно прикрепить к рубашке, или картинку, которую можно приклеить на знак остановки, чтобы обмануть ИИ-устройство (например, умную камеру наблюдения или беспилотный автомобиль) и заставить его думать, будто вы или знак – это любой другой объект на ваше усмотрение. Такой сценарий не выдумка, и он в действительности представляет угрозу для ИИ-систем на основе машинного обучения, а также для визуальных и аудиосистем, принимающих решения.

В 2017 г. исследователи из *Google* разработали круглые вставки с красочными психоделическими узорами, чтобы отвлечь классификаторы изображений с ИИ. Такие вставки могут обмануть ИИ-систему и заставить ее думать, что банан – или любой другой объект – это, например, тостер. Для этого достаточно просто поместить вставку рядом с объектом. Ранее с помощью других приемов экспериментаторы уже обманывали ИИ-системы, выдавая черепах за винтовки, а винтовки – за вертолеты. И хотя видоизмененные состязательные примеры со вставками хорошо заметны человеческому глазу, странные узоры и конструкции (скажем, граффити на стене здания или сложную трехмерную скульптуру) легко принять за обычное творчество – так что никто даже не догадается, что они призваны сбить с толку беспилотник, чтобы тот, например, принял больницу за военную цель.

В других экспериментах ИИ-системы неверно классифицировали знаки остановки как знаки ограничения скорости. В ряде работ уже анализировались потенциально опасные изменения, незаметные для человека, такие как подмена нескольких пикселей в изображении. В 2018 г. ученые из Калифорнийского университета в Беркли разработали состязательные примеры для систем распознавания речи. Иными словами, исследователи могли создать форму звуковой волны, почти идентичную реальной, и тем самым обмануть программу преобразования речи в текст, чтобы та превратила аудиосигнал в любую фразу по желанию исследователей.

Опыты с состязательным машинным обучением предполагают манипулирование тренировочными данными в ходе обучения ИИ. Хотя некоторые состязательные примеры можно нейтрализовать, потребовав от ИИ использовать сразу несколько систем классификаторов или запрограммировав их так, чтобы при обучении они не реагировали на состязательные примеры, во многих сферах применения ИИ сохраняются потенциальные риски.

СМ. ТАКЖЕ Смертоносные боевые роботы (1942), Машинное обучение (1959), Этика искусственного интеллекта (1976), Беспилотные автомобили (1984)

Исследования показали: если поместить круглые психоделические пятна в поле зрения ИИ-системы, можно обмануть ее и заставить «думать», что банан – это тостер. Это указывает на возможные риски при использовании ИИ в некоторых сферах.

Роботы собирают кубик Рубика. 2018

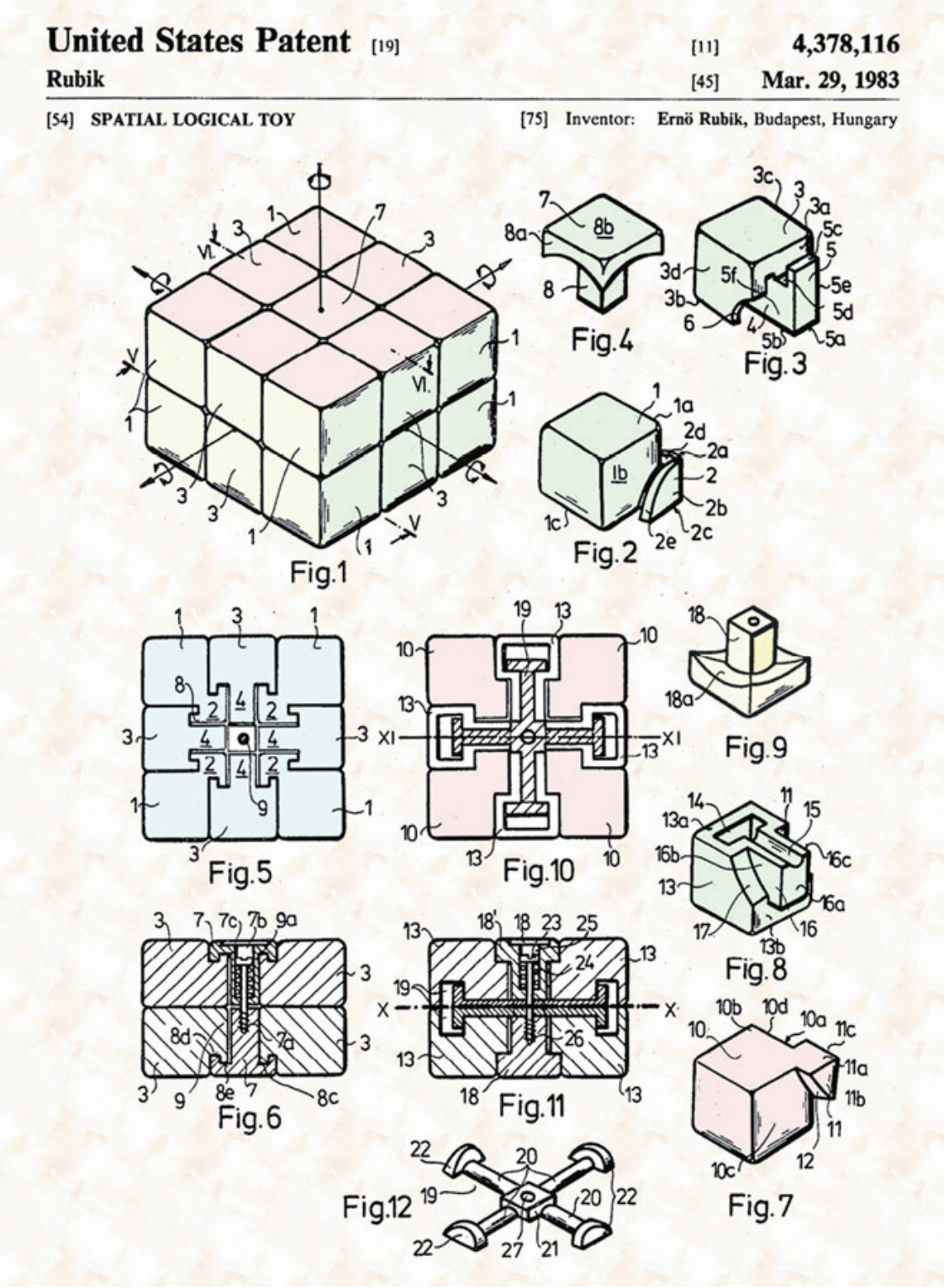
На протяжении многих лет специалисты по ИИ конструировали самых разных роботов, способных собирать кубик Рубика с помощью компьютерного зрения и физических манипуляций. Оригинальную головоломку в 1974 г. создал венгерский изобретатель Эрнё Рубик (р. 1944); к 1982 г. в Венгрии было продано 10 миллионов таких кубиков (как ни странно, это больше, чем население страны). К настоящему времени, согласно оценкам, в мире продано более 100 миллионов штук.

Кубик Рубика представляет собой конструкцию $3 \times 3 \times 3$ из меньших кубиков, окрашенных таким образом, что каждая из шести граней большого куба имеет свой цвет. Двадцать шесть внешних кубиков снабжены шарнирами, чтобы их можно было вращать. Цель в том, чтобы «собрать» кубик, то есть вернуть его в первоначальное состояние, в котором каждая грань состоит из квадратов одного цвета. Всего существует 43 252 003 274 489 856 000 вариантов расположения маленьких кубиков, и лишь один соответствует исходной позиции. Если бы у вас имелось по одному кубику для каждой возможной раскладки, вы могли бы покрыть ими всю поверхность Земли (включая океаны) примерно в 250 слоев.

В 2010 г. было доказано, что из любой конфигурации кубик можно собрать не более чем за двадцать ходов. В 2018 г. проворный робот под названием *Rubik's Contraption* преодолел полусекундный барьер, собрав кубик всего за 0,38 секунды (столько времени в сумме потребовалось на захват изображения, расчеты и физические манипуляции). В этом устройстве, созданном студентом-робототехником из Массачусетского технологического института Беном Кацем и разработчиком программного обеспечения Джаредом Ди Карло, использовались шесть двигателей *Kollmorgen ServoDisc* и так называемый двухфазный алгоритм Коцембы. (Для сравнения: в 2011 г. мировой рекорд среди роботов составлял 10,69 секунды.) В том же 2018 г. устройство, основанное на глубокой нейросети, научилось собирать кубик Рубика методом обучения с подкреплением без опоры на человеческие знания.

Существует вариация головоломки, которая никогда не появится на полках магазинов, – четырехмерная версия кубика Рубика, или тессеракт Рубика. Общее количество позиций тессеракта Рубика – головокружительная цифра: $1{,}76 \times 10^{120}$. Если бы вращающиеся грани тессеракта меняли положение каждую секунду с момента возникновения Вселенной, они и по сей день не успели бы продемонстрировать все возможные конфигурации.

СМ. ТАКЖЕ Ханойская башня (1883), Обучение с подкреплением (1951), Робот Шейки (1966), *ASIMO* и друзья (2000)



Патент США № 4378116, выданный Эрнё Рубику в 1983 г. на «пространственную логическую игрушку». На рисунке показано внутреннее устройство кубика.

Предсказатель смерти. 2019

В 2016 г. исследователи из Стэнфордского университета обучили ИИ-систему предсказывать, умрет ли человек в течение ближайших трех – двенадцати месяцев. Для данной книги этот результат важен как пример широкого спектра задач, которые в ближайшие годы будут решать ИИ и глубокое обучение.

Паллиативная помощь обычно предполагает облегчение боли, стресса и других симптомов пациента со смертельным диагнозом, когда выздоровление невозможно. Знание о том, когда потребуется такая помощь, облегчает жизнь самого пациента, его семьи и медперсонала, а также позволяет определить, в каких случаях паллиативный уход наиболее эффективен. Для создания ИИ под названием *Death Algorithm* группа ученых из Стэнфорда использовала данные примерно 170 тысяч пациентов, умерших от рака, сердечно-сосудистых, неврологических и других заболеваний. Информация из медицинских карт – включая диагноз, медицинские процедуры, результаты обследований, прописанные препараты и т. д. – использовалась в качестве входных данных для обучения ИИ-системы. Затем глубокая нейросеть тренировалась, корректируя веса нейронов. Нейросеть состояла из входного слоя размерности 13 654 (коды диагнозов и препаратов), 18 скрытых промежуточных слоев (каждый размерности 512) и выходного слоя скалярного типа.

В итоге девять из десяти человек, которым система предсказала смерть в ближайшие три – двенадцать месяцев, действительно умерли в течение этого срока. Кроме того, 95 % из тех, кому алгоритм напророчил более двенадцати месяцев жизни, действительно прожили дольше. Однако врач Сиддхартха Мукерджи в недавней статье в *New York Times* отмечает: «[Система глубокого обучения] учится, но не может объяснить нам, как именно; она назначает вероятности, но неспособна внятно их обосновать. Подобно ребенку, который осваивает велосипед методом проб и ошибок, а в ответ на просьбу сформулировать правила езды просто уносится вдаль, пожав плечами, алгоритм непонимающе смотрит на нас, когда мы спрашиваем: “Почему?” Это еще один черный ящик – как и сама смерть». И все же разработки предсказателей смерти на основе ИИ продолжаются. В 2019 г. группа экспертов из Ноттингемского университета показала, что машинное обучение превосходит по точности традиционные методы прогнозирования безвременной смерти, основанные на демографических, биометрических, клинических и бытовых факторах.

СМ. ТАКЖЕ Глубокое обучение (1965), Этика искусственного интеллекта (1976), Автономные роботы-хирурги (2016)

XIII
40 DEATH

Исследователи обучили ИИ-систему предсказывать, умрет ли человек в ближайшие 3–12 месяцев. А вы решились бы узнать день или хотя бы год своей смерти, будь у вас такая возможность?

Примечания и список литературы

> *ИИ – это следующий этап эволюции, но это уже иной этап… Одно устройство с ИИ может рассказать другому не только все, что знает, – в том смысле, в котором учитель передает свои знания ученику, – оно также может сообщить ему все о собственной конструкции… В сущности, человеческий разум имеет мало общего с Богом или компьютером. Он гораздо больше похож на разум шимпанзе и [рассчитан] на жизнь в джунглях и полях.*
> ***Эдвард Фредкин. Цитируется по книге Памелы Маккордак «Машины, которые думают»***

Я составил список литературы, в котором привел некоторые материалы, использованные мною для исследований и написания этой книги – в том числе для поиска цитат. Как известно многим читателям, интернет-сайты живут не вечно: иногда они меняют адреса или исчезают. По указанным здесь ссылкам можно было найти полезную справочную информацию, когда писалась эта книга. Такие интернет-источники, как Википедия, – удобные отправные точки для любых изысканий, и иногда я использовал этот сайт в качестве стартовой площадки, наряду со многими другими ресурсами, книгами и научными статьями.

Если я упустил из виду интересный или поворотный момент в истории ИИ, который, на ваш взгляд, заслуживает внимания, пожалуйста, дайте мне об этом знать. Просто зайдите на мой сайт (*pickover.com*) и напишите мне электронное письмо с объяснением этой идеи и того, как она, по вашему мнению, изменила мир. Возможно, в последующие издания этой книги войдут полностраничные заметки о таких феноменах, как Василиск Роко, порождающие состязательные сети, нейроморфные вычисления, байесовские сети, телесериал «Мир Дикого Запада», кинофильм 1983 г. «Военные игры», *LSTM*-сети (сети с долгой краткосрочной памятью), и о многом другом. Наконец, я хочу поблагодарить моих редакторов Мередит Хейл и Джона Мейлса, а также Денниса Гордона, Тома Эриксона, Майкла Перроне, Тею Красек и Пола Московица за их комментарии и предложения.

Ознакомительное чтение

Crevier, D., *AI* (New York: Basic Books, 1993).

Dormehl, L., *Thinking Machines* (New York: Tarcher, 2017).

McCorduck, P., *Machines Who Think* (Natick, MA: A. K. Peters, 2004).

Nilsson, N., *The Quest for Artificial Intelligence* (New York: Cambridge University Press, 2010).

Riskin, J., *The Restless Clock* (Chicago: University of Chicago Press, 2016).

Truitt, E., *Medieval Robots* (Philadelphia: University of Pennsylvania Press, 2015).

Walsh, T., *Machines That Think* (London: C. Hurst & Co., 2017).

Библиография

Предисловие

Hambling, D., “Lethal logic,” *New Scientist,* vol. 236, no. 3151, p. 22, Nov. 11–17, 2017.

Reese, M., “Aliens, Very Strange Universes and Brexit – Martin Rees Q&A,” *The Conversation*, April 3, 2017, http://tinyurl.com/mg3w6ez

Truitt, E., *Medieval Robots* (Philadelphia: University of Pennsylvania Press, 2015).

“Visual Trick Has AI Mistake Turtle for Gun,” *New Scientist*, vol. 236, no. 3151, p. 19, November 11–17, 2017.

Ок. 400 до н. э. Талос

Haughton, B., *Hidden History: Lost Civilizations, Secret Knowledge, and Ancient Mysteries* (Franklin Lakes, NJ: New Page Books, 2007).

Ок. 250 до н. э. Водяные часы Ктесибия

Dormehl, L., *Thinking Machines* (New York: Tarcher, 2017).

Ок. 190 до н. э. Счеты

Ewalt, D., “No. 2 The Abacus,” *Forbes*, August 30, 2005, http://tinyurl.com/yabaocr5

Krimmel, J., “Artificial Intelligence Started with the Calendar and Abacus,” 2017, https://tinyurl.com/y5tnoxbl

Ок. 125 до н. э. Антикитерский механизм

Garnham, A., *Artificial Intelligence: An Introduction* (London: Routledge, 1988).

Marchant, J., “The Antikythera Mechanism: Quest to Decode the Secret of the 2,000-Year-Old Computer,” March 11, 2009, http://tinyurl.com/ca8ory

1206. Автоматы Аль-Джазари

Hill, D., *Studies in Medieval Islamic Technology*, ed. D. A. King (Aldershot, Great Britain: Ashgate, 1998).

Ок. 1220. Медные рыцари из легенды о Ланселоте

Riskin, J., *The Restless Clock* (Chicago: University of Chicago Press, 2016).

Truitt, E., *Medieval Robots* (Philadelphia: University of Pennsylvania Press, 2015).

Ок. 1300. Механический парк в Эдене

Bedini, S., “The Role of Automata in the History of Technology,” *Technology and Culture*, vol. 5, no. 1, p. 24–42, 1964.

Lightsey, S., *Manmade Marvels in Medieval Cultures and Literature* (New York: Palgrave, 2007).

Ок. 1305. *Ars Magna* Раймунда Луллия

Dalakov, G., “Ramon Llull,” http://tinyurl.com/ybp8rz28

Gray, J., “‘Let us Calculate!’: Leibniz, Llull, and the Computational Imagination,” http://tinyurl.com/h2xjn7j

Gardner, M., *Logic Machines and Diagrams* (New York: McGraw-Hill, 1958).

Madej, K., *Interactivity, Collaboration, and Authoring in Social Media* (New York: Springer, 2016).

Nilsson, N., *The Quest for Artificial Intelligence* (New York: Cambridge University Press, 2010).

1352. Религиозные автоматоны

Coe, F., *The World and Its People, Book V, Modern Europe*, ed. L. Dunton (New York: Silver, Burdett, 1896).

Fraser, J., *Time, the Familiar Stranger* (Amherst: University of Massachusetts Press, 1987).

Ок. 1495. Робот-рыцарь Леонардо да Винчи

Phillips, C., and S. Priwer, *The Everything Da Vinci Book* (Avon, MA: Adams Media, 2006).

Rosheim, M., *Leonardo's Lost Robots* (New York: Springer, 2006).

1580. Голем

Blech, B., "Stephen Hawking's Worst Nightmare? Golem 2.0" (tagline), *The Forward*, January 4, 2015, http://tinyurl.com/yats534k

1651. «Левиафан» Гоббса

Dyson, G., *Darwin among the Machines* (New York: Basic Books, 1997).

1714. Мельница Лейбница

Bostrom, N., "The Simulation Argument: Why the Probability that You Are Living in a Matrix is Quite High." *Times Higher Education Supplement*, May 16, 2003, http://tinyurl.com/y8qorjcf

Moravec, H., "Robot Children of the Mind." In David Jay Brown's *Conversations on the Edge of the Apocalypse* (New York: Palgrave, 2005).

1726. Станок для написания книг в Лагадо

Weiss, E., "Jonathan Swift's Computing Invention." *Annals of the History of Computing*, vol. 7, no. 2, p. 164–165, 1985.

1738. Механическая утка де Вокансона

Glimcher, P., *Decisions, Uncertainty, and the Brain: The Science of Neuroeconomics.* (Cambridge, MA: MIT Press, 2003).

Riskin, J., "The Defecating Duck, or, the Ambiguous Origins of Artificial Life." *Critical Inquiry*, vol. 29, no. 4, p. 599–633, 2003.

1770. Механический турок

Morton, E., "Object of Intrigue: The Turk, a Mechanical Chess Player that Unsettled the World." August 18, 2015, http://tinyurl.com/y72aqfep

1774. Автоматоны Жаке-Дро

Lorrain, J., *Monsieur De Phocas* (trans. F. Amery) (Sawtry, Cambridgeshire, UK: Dedalus, 1994).

Riskin, J., *The Restless Clock* (Chicago: University of Chicago Press, 2016).

1818. «Франкенштейн»

D'Addario, D., "The Artificial Intelligence Gap Is Getting Narrower," *Time*, October 10, 2017, http://tinyurl.com/y8g5bu5o

Gallo, P., "Are We Creating a New Frankenstein?" *Forbes*, March 17, 2017, http://tinyurl.com/ycsdr6gt

1821. Машинное творчество

Colton, S., and G. Wiggins, "Computational Creativity: The Final Frontier?" In *Proceedings of the 20th European Conference on Artificial Intelligence,* 2012.

1854. Булева алгебра

Titcomb, J., "Who is George Boole and Why is He Important?" *The Telegraph*, November 2, 2015, http://tinyurl.com/yb25t8ft

1863. «Дарвин среди машин»

Wiener, N., "The Machine Age," 1949 unpublished essay for the *New York Times*, http://tinyurl.com/ybbpeydo

1868. «Паровой человек в прериях»

Liptak, A., "Edward Ellis and the Steam Man of the Prairie," *Kirkus*, November 6, 2015, http://tinyurl.com/yadhxn7t

1907. Тик-Ток

Abrahm, P., and S. Kenter, "Tik-Tok and the Three Laws of Robotics," *Science Fiction Studies*, vol. 5, pt. 1, March 1978, http://tinyurl.com/ybm6qv2y

Goody, A., *Technology, Literature and Culture* (Malden, MA: Polity Press, 2011).

1920. *R.U.R.*

Floridi, L., *Philosophy and Computing* (New York: Taylor & Francis, 2002).

Stefoff, R., *Robots* (Tarrytown, NY: Marshall Cavendish Benchmark, 2008).

1927. «Метрополис»

Lombardo, T., *Contemporary Futurist Thought* (Bloomington, IN: AuthorHouse, 2008.)

1942. Три закона робототехники Азимова

Markoff, J., "Technology: A Celebration of Isaac Asimov," *New York Times*, April 12, 1992, http://tinyurl.com/y9gevq6t

1943. Искусственные нейронные сети

Lewis-Kraus, G., "The Great A.I. Awakening," *New York Times Magazine*, December 14, 2016, http://tinyurl.com/gue4pdh

1950. «Человеческое использование человеческих существ»

Crevier, D., AI (New York: Basic Books, 1993).

Wiener, N., *The Human Use of Human Beings* (London: Eyre & Spottiswoode, 1950).

1952. Распознавание речи

"Now We're Talking: How Voice Technology is Transforming Computing," *The Economist*, January 7, 2017, http://tinyurl.com/yaedcvfg

1954. Обработка естественного языка

“701 Translator,” IBM Press Release, January 8, 1954, http://tinyurl.com/y7lwblng

1956. Дартмутский семинар по искусственному интеллекту

Dormehl, L., *Thinking Machines* (New York: Tarcher, 2017).

McCorduck, P., *Machines Who Think* (Natick, MA: A. K. Peters, 2004).

1957. Трансгуманизм

Huxley, J., *New Bottles for New Wine* (London: Chatto & Windus, 1957).

Istvan, Z., “The Morality of Artificial Intelligence and the Three Laws of Transhumanism,” *Huffington Post*, http://tinyurl.com/ycpx9bwa

Pickover, C., A *Beginner’s Guide to Immortality* (New York: Thunder’s Mouth Press, 2007).

1959. Представление знаний

Nilsson, N., The *Quest for Artificial Intelligence* (New York: Cambridge University Press, 2010).

1964. Психотерапевт *ELIZA*

Weizenbaum, J., “ELIZA – A Computer Program for the Study of Natural Language Communication Between Man and Machine,” *Communications of the ACM*, vol. 9, no. 1, p. 36–45, 1966.

1964. Распознавание лиц

West, J., “A Brief History of Face Recognition,” http://tinyurl.com/y8wdqsbd

1965. Экспертные системы

Dormehl, L., *Thinking Machines* (New York: Tarcher, 2017).

1965. Нечеткая логика

Carter, J., “What is ‘Fuzzy Logic’?” *Scientific American*, http://tinyurl.com/yd24gngp

1965. Глубокое обучение

Fain, J., “A Primer on Deep Learning,” *Forbes*, December 18, 2017, http://tinyurl.com/ybwt9qp3

1967. Жизнь в симуляции

Davies, P., “A Brief History of the Multiverse,” *New York Times*, 2003, http://tinyurl.com/y8fodeoy.

Koebler, J., “Is the Universe a Giant Computer Simulation?” http://tinyurl.com/y9lluy7a

Reese, M., “In the Matrix,” http://tinyurl.com/y9h6fjyx.

1972. Параноик *PARRY*

Wilks, Y., and R. Catizone, “Human-Computer Conversation,” arXiv: cs/9906027, June 1999, http://tinyurl.com/y7erxtxm

1975. Генетические алгоритмы

Copeland, J., *The Essential Turing* (New York: Oxford University Press, 2004).

Dormehl, L., *Thinking Machines* (New York: Tarcher, 2017).

1979. Победа над чемпионом по коротким нардам

Crevier, D., AI (New York: Basic Books, 1993).

1982. «Бегущий по лезвию»

Guga, J., "Cyborg Tales: The Reinvention of the Human in the Information Age," in *Beyond Artificial Intelligence* (New York: Springer, 2015).

Littman, G., "What's Wrong with Building Replicants?" in *The Culture and Philosophy of Ridley Scott* (Lanham, MD: Lexington).

1984. Беспилотные автомобили

Lipson, H., and M. Kurman, *Driverless* (Cambridge, MA: MIT Press, 2016).

1986. Роевой интеллект

Jonas, David, and Doris Jonas, *Other Senses, Other Worlds* (New York: Stein and Day, 1976).

1988. Парадокс Моравека

Elliott, L., "Robots Will Not Lead to Fewer Jobs – But the Hollowing Out of the Middle Class." *The Guardian*, August 20, 2017, http://tinyurl.com/y7dnhtpt

Moravec, H., *Mind Children* (Cambridge, MA: Harvard University Press, 1988).

Pinker, S., *The Language Instinct* (New York: William Morrow, 1994).

1990. «Слоны не играют в шахматы»

Brooks, R., "Elephants Don't Play Chess," *Robotics and Autonomous Systems*, vol. 6, p. 139–159, 1990.

Shasha, D., and C. Lazere, *Natural Computing* (New York: Norton, 2010).

1993. Герметичный ящик для искусственного интеллекта

Возможно, читателям будет интересно познакомиться с Василиском Роко – мысленным экспериментом, в котором ИИ-системы из будущего мстят людям, не создававшим ИИ-системы. Во многих версиях мысленного эксперимента эти системы задним числом наказывают людей, подвергая их пыткам.

Vinge, V., "The Coming Technological Singularity." VISION-21 Symposium, March 30–31, 1993.

1994. Шашки и искусственный интеллект

Madrigal, A., "How Checkers Was Solved." *The Atlantic*, July 19, 2017, http://tinyurl.com/y9pf9nyd

1997. *Deep Blue* обыгрывает чемпиона мира по шахматам

Webermay, B., "Swift and Slashing, Computer Topples Kasparov." May 12, 1997, http://tinyurl.com/yckh6xko

2000. *ASIMO* и друзья

Wiener, N., *God and Golem* (Cambridge, MA: MIT Press, 1964).

2001. «Искусственный разум» Спилберга

Gordon, A., *Empire of Dreams: The Science Fiction and Fantasy Films of Steven Spielberg* (New York: Rowman & Littlefield, 2007).

2002. Робот-пылесос *Roomba*

Reel, M., “How the Roomba Was Realized.” Bloomberg, October 6, 2003, http://tinyurl.com/yd4epat4

2002. Решение для игры вари

Romein, J., and H. Bal, “Awari is Solved.” *ICGA Journal*, September 2002, p. 162–165.

2003. Максимизатор скрепок

Следует отметить, что специалист по ИИ Элиезер Юдковский (р. 1979) говорит, что идея максимизатора скрепок, возможно, впервые возникла именно у него. См. подкаст “Waking Up with Sam Harris #116–AI: Racing Toward the Brink” (с Элиезером Юдковским).

2006. *Quackle* выигрывает в скребл

Anderson, M., “Data Mining Scrabble.” *IEEE Spectrum*, vol. 49, no. 1, p. 80.

2011. *Watson* участвует в телевикторине

Jennings, K., “My Puny Human Brain.” *Slate*, Feb. 16, 2011, http://tinyurl.com/86xbqfq

2015. «Назовем их искусственными пришельцами»

Kelly, K., “Call them Artificial Aliens,” in Brockman, J., ed., *What to Think About Machines That Think* (New York: Harper, 2015).

2015. Искусственный интеллект на Марсе

Fecht, S., “The Curiosity Rover and Other Spacecraft Are Learning to Think for Themselves.” *Popular Science*, June 21, 2017, http://tinyurl.com/y895pq6k,

Koren, M., “The Mars Robot Making Decisions On Its Own,” *The Atlantic*, June 23, 2017, http://tinyurl.com/y8s8alz6

2016. *AlphaGo* – чемпион по игре в го

Chan, D., “The AI That has Nothing to Learn from Humans.” *The Atlantic*, October 20, 2017. http://tinyurl.com/y7ucmuzo

Ito, J., and J. How, *Whiplash: How to Survive Our Faster Future* (New York: Grand Central Publishing, 2016).

2018. Состязательные примеры

Brown, T., et al., “Adversarial Patch,” 31st Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, 2017.

2019. Предсказатель смерти

Avati, A., et al., “Improving Palliative Care with Deep Learning,” *IEEE International Conference on Bioinformatics and Biomedicine* (BIBM), Kansas City, MO, p. 311–316, 2017.

Mukherjee, S., “This Cat Sensed Death. What if Computers Could, Too?” New York Times, January 3, 2018, http://tinyurl.com/yajko6pv

Rajkomar, A., et al., “Scalable and accurate deep learning with electronic health records,” npj Digital Medicine, vol. 1, no. 18, 2018, http://tinyurl.com/ych74oe5

Иллюстрации

Некоторые из использованных в книге старых и редких иллюстраций было трудно найти в хорошем качестве, не нарушив при этом авторских прав, поэтому в ряде случаев я взял на себя смелость обработать изображения. Я убрал пятна и царапины, восстановил выцветшие фрагменты, а кое-где добавил контраста черно-белым элементам ради большей выразительности. Надеюсь, историки-пуристы простят мне эту небольшую художественную вольность и отнесутся к ней с пониманием. Мною двигало лишь одно желание: чтобы книга получилась красивой и заинтересовала самую широкую аудиторию, а приведенные в ней фотографии и рисунки в полной мере передавали мое восхищение перед удивительным и многообразным миром искусственного интеллекта.

Alamy: The Advertising Archives; David Fettes; richterfoto; Science Photo Library

Courtesy of International Business Machines Corporation, © (1962) International Business Machines Corporation

Courtesy of Universal Studios Licensing LLC: endpapers

Cyclopaedia: Abraham Rees

Gallica: Robert de Boron; Guillaume de Machaut

Internet Archive: J.J. Grandville; Ramon Llull

iStock/Getty Images Plus: ChubarovY; in-future; Kickimages; Paulbr; Vera Petruk; PhonlamaiPhoto; JIRAROJ PRADITCHAROENKUL; Rouzes; sergeyryzhov; Soifer; Vladimir Timofeev

Getty: Al Fenn; Kyodo News/Contributor; The Washington Post/Contributor; Westend61

Google Research Team

Library of Congress: Charles Verschuuren

New York World's Fair 1939-1940 records, Manuscripts and Archives Division, The New York Public Library/Mansfield Memorial Museum: endpapers

Scientific American

Shutterstock: 7th Son Studio; Charles Adams; Berke; Black Moon; gualtiero boffi; Willyam Bradberry: endpapers; camilla$$; Chesky; Esteban De Armas; Digital Storm; Dmitry Elagin; Leonid Eremeychuk; Evannovostro; Everett Historical; Ilterriorm; Eugene Ivanov: back cover; Ala Khviasechka; Anastasiia Kucherenko; MicroOne; Morphart Creation; MossStudio; NadyaEugene; Nor Gal; Ociacia: back cover; Phonlamai Photo; Photobank gallery; PHOTOCREO/ Michal Bednarek: endpapers; Saran_Poroong; Glenn Price; Quality Stock Arts; R-Type; ra2studio; Robsonphoto; Benjavisa Ruangvaree; Tatiana Shepeleva; Travel mania; Trial; Tsuneomp; videodoctor

Stanford University: Chuck Painter

United States Patent and Trademark Office

Wikimedia: Rocky Acosta; Ismail al-Jazari/MFA Boston; Aquarius70; Edward S. Ellis/Cover of "Steam Man of the Prairies"; Emesee/Electric Sheep; David Iliff; Jastrow/Ludovisi Collection; Erik Möller, Mensch-Erfinder-Genie exhibit, Berlin; Jessica Mullen; NASA: back cover; John R. Neill; New York World's Fair 1939; Joseph Racknitz/Humboldt University Library; Rama; William Bruce Ellis Ranken/Christie's; SRI International; Sybil Tawse/"Bulfinch's Mythology"; Robert Toombs/New York Five Cent Library; Carston Ullrich: back cover; Mogi Vicentini; Theodore Von Holst; Jiuguang Wang

Об авторе

Клиффорд Пиковер (р. 1957) – ученый, изобретатель, популяризатор науки, автор более 50 книг по самым разным темам, от математики, физики и биологии до истории, искусства и религии. Обладатель степени PhD Йельского университета; его изобретениями (более 700 патентов) пользуются компании *IBM, Google, eBay, Yahoo! PayPal, LinkedIn* и другие. У него больше 40 тысяч подписчиков в Twitter, а сайт *Pickover.com* посещают миллионы интернет-пользователей.

> *«Пиковер видит миры, лежащие за пределами той реальности, что известна нам».*
> *The New York Times*

> *Артур Кларк мыслил масштабно, Баки Фуллер мыслил масштабно, но Клифф Пиковер превосходит их обоих.*
> *WIRED*

> *…Две части Айзека Азимова, по одной части Карла Сагана и Мартина Гарднера – получается Клиффорд Пиковер, один из самых ярких и заставляющих думать авторов современности.*
> *SKEPTIC*